University of Alberta

ACCOUNTING FOR GEOLOGICAL BOUNDARIES IN GEOSTATISTICAL MODELING OF MULTIPLE ROCK TYPES

by

Paula Larrondo ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Master of Science**

in

Mining Engineering

Department of Civil and Environmental Engineering

Edmonton, Alberta

Fall 2004

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-95790-X
Our file Notre référence
ISBN: 0-612-95790-X

The author has granted a non-exclusive license allowing the Library and Archives Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

As far as the laws of mathematics refer to reality, they are not certain;
and as far as they are certain, they do not refer to reality

Albert Einstein

To
Julia
and
Lucy

Acknowledgements

When I first thought of pursuing formal studies in Geostatistics it saw it more as a dream than a possibility. For making that dream come true I have to thank Dr. Clayton Deutsch. Not only has he made this possible in terms of funding and patient guidance, but he has also generously shared knowledge and bright ideas. These years under his guidance have intensified my research spirit.

Financial support for this research was provided by industry sponsors of the Centre for Computational Geostatistics (CCG) and the University of Alberta; their support is greatly appreciated.

I also would like to thank the graduate students and members of the CCG for their companionship along the way. I would especially like to thank Oy Leuangthong, Willie and Cathy Hamilton, and Julián Ortiz for their friendship and for making me feel at home away from home.

A special thank you goes to my first boss Marko Dydik for encouraging me to study geostatistics. And to Maritza Alarcón, for change my view of mathematics back when I was a high school student.

Finally and most importantly, I would like to thank my mother Lucy for her love and unconditional support. I would not be able to do this without her. And to my daughter Julia, her smiles mean the world to me.

Table of Contents

1 Introduction **1**

Background 1

Proposed Methodology 4

Thesis Outline 8

2 Global Linear Model of Coregionalization **9**

Theoretical Background 9

Illustration of Theory 13

2D Example 19

Application 23

Variography 24

Simulation 28

Validation 29

Comparison at the boundary 35

3 Local Non-stationary Model of Coregionalization **37**

Theoretical Background 37

Maximum Distance of Influence 38

Precedence Rules 40

Statistical Parameters 41

Mean 43

Variance 43

Covariance 44

Optimization of the Statistical Parameters 47

Mean Optimization 48
Variance Optimization 50
Covariance Optimization 51
Estimation in presence of local non-stationary boundaries 52

4 Implementation Details 54
RT model and boundaries 54
Mean optimization 59
Variance optimization 61
Covariance optimization 62
Non-stationary kriging within boundaries 66
1-D Example 70

5 Application 81
Data generation 81
Estimation 86

6 Conclusions and Future Work 98
Conclusions 98
Future Work 100

References 102

List of Figures

Figure 1.1: Grade profile of two drill holes in a porphyry copper deposit in Northern Chile. 3

Figure 1.2: Schematic cross-section illustrating porphyry copper mineralisation. 5

Figure 1.3: Application of a local model of coregionalization. 6

Figure 1.4: Local non-stationary soft boundary between rock type 1 and 2. 6

Figure 1.5: Difference between a global soft boundary and a local non-stationary boundary. 7

Figure 1.6: Schematic cross-section of a structural boundary between two porphyries. 8

Figure 2.1: Example of two domains and the corresponding categorical models. 15

Figure 2.2: Cross-covariance reproduction of the simulated random variables Z_1 and Z_2, assuming both variables are collocated. 16

Figure 2.3: (A) Cross-covariance between Z_1 and Z_2 combined side by side. (B) Cross-covariance between Z_1 and Z_2 combined using ellipsim categorical model as a boundary model with a target proportion of Z_1 of 50%. 16

Figure 2.4: Cross-covariance between Z_1 and Z_2 combined using ellipsim categorical model as a boundary model, for target proportion of Z_1 of 5, 10, 20 and 50%. 17

Figure 2.5: Cross-covariance between Z_1 and Z_2 combined using ellipsim categorical model as a boundary model for different radii: 25, 50, 100 and 200 m. 18

Figure 2.6: Categorical rock type model. 20

Figure 2.7: A realization of the merged grades Z_1 up to Z_5 using the categorical rock type model of Figure 2.6. 20

Figure 2.8: Cross-covariance reproduction of the simulated pairs Z_i and Z_j for $i \neq j$, combined by the categorical rock type model. . 22

Figure 2.9: Reference or 'true' image. Z_1 is represented by $\mathbf{RT_1}$, Z_2 by $\mathbf{RT_2}$. 24

Figure 2.10: Sketch with the structured cross covariance and calculated sill of a cross variogram given an experimental cross covariance between two non-collocated variables. 25

Figure 2.11: Direct and cross variograms and LMC model for Z_1 and Z_2. 27

Figure 2.12: Variogram model for Z_1 and Z_2, obtained independently. 28

Figure 2.13: Direct variograms reproduction for Z_1 and Z_2, cosimulated and independently simulated. 30

Figure 2.14: Cross covariance reproduction for Z_1 and Z_2, cosimulated and independently simulated. 30

Figure 2.15: Accuracy plot for Z_1 and Z_2, estimate independently versus cosimulated. 31

Figure 2.16: Cross validation of data values in $\mathbf{RT_1}$ and $\mathbf{RT_2}$, estimate independently versus cosimulated. 32

Figure 2.17: Error (true-estimated) distribution for Z_1 and Z_2, estimate independently versus cosimulated. 33

Figure 2.18: Global mean and variance reproduction for Z_1 and Z_2; cosimulation and independent simulation. 34

Figure 2.19: Correlation coefficient between E-type estimates of cosimulated and independently simulated models. 35

Figure 2.20: Average variance calculated from blocks within a given distance from the boundary between Z_1 and Z_2. 36

Figure 3.1: Maximum distances of influence for the boundary zone between rock type k and rock type p. 39

Figure 3.2: Boundary zone $\mathbf{BZ_{pk}}$ and $\mathbf{BZ_{kp}}$ corresponding to the area of influence of rock type p into rock type k and vice versa. 39

Figure 3.3: Three rock type example where a predefined precedence rule is used to determine the precedent rock type. 40

Figure 3.4: Three-rock type example where a neutral arrangement. 41

Figure 3.5: Decomposition of the random function $Z(\mathbf{u})$ in two stationary variables Z_k and Z_p, and a non-stationary boundary variable $Z_{kp}(\mathbf{u})$. 42

Figure 3.6: Two-rock type example showing the stationary and non-stationary components of the covariance model. 48

Figure 3.7: Mean and variance of the random variable Z_{kp} $(\mathbf{u})$. 49

Figure 3.8: Non-stationary covariance of $Z_{kp}(\mathbf{u})$. 51

Figure 4.1: Flow chart of the program `boundmod`. 55

Figure 4.2: Parameter file for `boundmod` program. 56

Figure 4.3: Example of a geological model with 4 rock types and maximum distances of influence define for each boundary. 57

Figure 4.4: Set of four precedence rules for the example of Figure 4.3. 58

Figure 4.5: Output of the `boundmod` program for the example of Figure 4.3, considering a set of 4 precedence rules. 58

Figure 4.6: Output of the `boundmod` program for the example of Figure 4.3, considering a neutral arrangement. 58

Figure 4.7: Flow chart of the program `opt_mean`. 60

Figure 4.8: Parameter file for `opt_mean` program. 61

Figure 4.9: Parameter file for `opt_var` program. 62

Figure 4.10: Flow chart of the program `opt_var`. 63

Figure 4.11: Parameter file for `opt_cov` program. 64

Figure 4.12: Flow chart of the program `opt_cov`. 65

Figure 4.13: Flow chart of the program `kt3d_bound`. 67

Figure 4.14: Flow chart for the filling of the kriging matrix and right hand side of kriging system in `kt3d_bound` program. 68

Figure 4.15: Parameter file for `kt3d_bound` program. 69

Figure 4.16: Linear functions for the mean and variance used to transform the normal score values of **SGS3**. 71

Figure 4.17: Scheme of how the three unconditional simulations are add to obtain $Z(\mathbf{u})$. 71

Figure 4.18: Dataset for 1-D example. 72

Figure 4.19: Optimized mean obtained for dataset of 1-D example. 73

Figure 4.20: Output file of `opt_mean` program for the 1-D example. 74

Figure 4.21: Optimized variance obtained for 1-D example. 74
Figure 4.22: Output file of `opt_var` program for the 1-D example. 75
Figure 4.23: Experimental covariance from pairs within the boundary zone, optimum non-stationary covariance obtained from `opt_cov` and original covariance of the non-stationary component. 75
Figure 4.24: Output file of `opt_cov` program for the 1-D example. 76
Figure 4.25: Optimization convergence of mean, variance and covariance parameters. 76
Figure 4.26: Data configuration for the estimation of an unknown location with the 1-D example. 77
Figure 4.27: Grade reproduction profile along the X-coordinate. Reference values versus kriging estimates. 79
Figure 4.28: Scatter plot reference values versus kriging estimates, for the 1-D example. 80
Figure 4.29: Reference values versus kriging estimates considering 2000 nodes instead of 200. 80

Figure 5.1: Categorical rock type model of a porphyry copper deposit in Northern Chile. 82
Figure 5.2: Boundary zone and distance to boundary maps for the same section and level of the rock type model of Figure 5.1. 84
Figure 5.3: Transformed variable $Y_6(\mathbf{u})$. 85
Figure 5.4: Section and Level maps of the reference distribution. 85
Figure 5.5: Output file of `opt_mean` program. 86
Figure 5.6: Output file of `opt_var` program. 87
Figure 5.7: Output file of `opt_cov` program. 88
Figure 5.8: Optimization convergence of mean and variance in terms of the objective function value. 89
Figure 5.9: Covariance optimization convergence in terms of the range for the 6 boundary zones. 90
Figure 5.10: Scatter plot of the reference versus the estimate for each boundary zone. 92
Figure 5.11: Scatter plot of the reference versus the estimate within the stationary portions of each rock type. 93

Figure 5.12: Mean at the non-stationary boundary zone. 94
Figure 5.13: Variance at the non-stationary boundary zone. 95
Figure 5.14: Cross validation results. 96
Figure 5.15: Cross validation results comparison with ordinary kriging at the boundary zone. 97

List of Tables

Table 2.1: LMC model for Z_1 and Z_2. 26

Table 2.2: Variogram models for Z_1 and Z_2, obtained considering the two variables independently. 26

Table 2.3: Parameters used for cosimulation of Z_1 and Z_2. 28

Table 2.4: Parameters used for independent simulations of Z_1 and Z_2. 29

Table 5.1: Comparison between the reference and optimum a_{kp} and b_{kp}. 87

Table 5.2: Variogram models for the stationary regions of rock type 2 to 5. 88

Table 5.3: Comparison of the mean and variance of the each rock type (stationary regions) between the reference training image and the kriging results. 91

Chapter 1

Introduction

There are multiple rock types in virtually all mineral deposits. For example, a porphyry intrusive within some host rock. There may also be multiple rock types within the porphyry that are dominated by different styles of mineralisation. The contact between rock types is often gradational and not abrupt. In fact, the contact or boundary is often a zone where there are influences from the adjacent rock types and from the zone itself. Fractures or other geomechanical or geochemical changes at the boundary can cause enrichment. The size of these boundary zones can be significant. They can be tens of meters, which has an impact on mining and mineral resources appraisal.

Classical geostatistical modeling assumes that there is no change in the average grade and variability over the domain. Therefore it is often assume that the contacts between rock types are abrupt and data is separated into the different rock types. Another approach is to divide data into three zones, one within each rock type and separate boundary zone as a third. The latest usually has too few data and a more complex behavior that could include trends, correlation across the boundary among others features.

This thesis is concerned with techniques to more reasonably model the grades in and near boundary zones between rock types. A number of techniques and statistical models will be developed for this purpose.

Background

Mineral resource and ore reserve estimation requires a critical decision regarding the geological domains that will be used for the grade modeling, as well as the type of boundaries between these domains. The statistical characteristics of the domains can

have a significant impact on the final mineral inventory. The most common geostatistical techniques, such as kriging and sequential simulation, are based on strong assumptions of stationarity of the estimation domains. In particular, they are based on a second order stationary hypothesis, that is, the mean, variance and covariance remain constant across the entire domain; they do not depend on the location of the support points but only on the distance between them. Therefore, geological domains should be chosen as statistically homogeneous zones, which are geologically significant and coherent, but with enough data to allow reliable inference of the required statistics.

Wherever possible, geology should be used to define appropriate estimation domains. Some deposits will have a very simple geometry and mineralisation pattern, making the choice of estimation domains straightforward from geological units. In other cases the mineralisation of interest is not entirely defined by a single geological unit or may transgress their boundaries. In these deposits, it is common to find an important structural or lithological influence on grade distribution and multiple events of mineralisation. In these cases, a combination of different geological attributes may reflect an estimation domain. When it is not possible to identify and isolate the geological controls of grade distribution, the estimation domains can be defined purely by grade boundaries; however, this method is considerably more risky in that under or overestimation of grade and tonnage can result.

Once certain geological or grade boundaries are selected to represent an estimation boundary, the choice should be validated statistically. Differences in the mean, variance and covariance between domains are common. Complex features on a probability plot or a high coefficient of variation may indicate that the chosen domain still has mixed populations. Indicator variograms can be used to test for consistency in anisotropy and grade continuity of different grade ranges within a proposed domain.

Following the definition of the estimation domains, an analysis on how grades change across the boundaries between domains should be done. This validates the proposed units and determines the nature of their boundaries. Domain boundaries are often referred to as either 'hard' or 'soft'. Hard boundaries are found when an abrupt change in average grade or variability occurs at the contact between two domains, such as coal seams or sedimentary zinc deposits. In deposits where the disseminated mineralisation has a gradational nature, such as some porphyry copper deposits, and grades change transitionally across a boundary, the contact is referred to as a soft boundary (Figure 1.1). Other examples are nickel, chromium and platinum orthomagmatic deposits that often show an increasing grade profile towards the bottom of a layer due to magmatic segregation.

Hydrothermal deposits with successive injections of circulating fluids produce concentric fracturing and trends in the mineralisation.

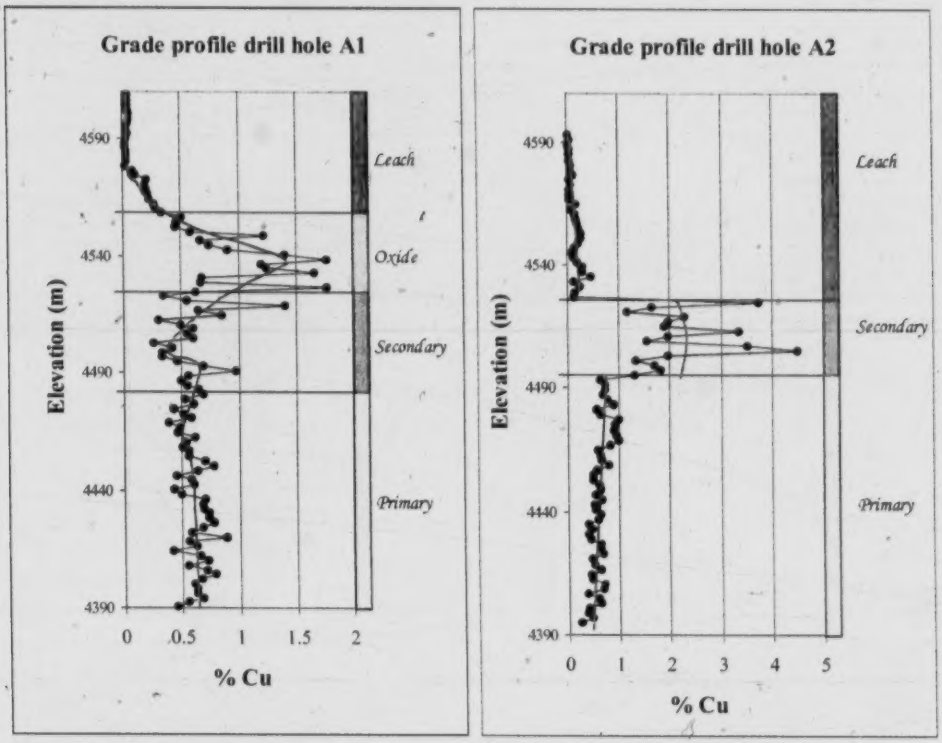

Figure 1.1: Grade profile of two drill holes in a porphyry copper deposit in Northern Chile, showing an example of soft boundaries (left) between mineralisation units as well as an example of hard boundaries (right). Notice how the mean and variance change across the boundaries.

Correct representation of soft boundaries should ensure the reproduction of the correlation of the grades across the boundary and ensure reproduction of non-stationary variations of the mean, variance and covariance in the zone of influence of each rock type. Improved modelling of boundaries would benefit long and medium term mine planning by reducing the percentage of misclassified blocks, obtaining more accurate tonnage, average grade and metal content estimates. The estimation of dilution for underground and open pit deposits would also be improved. When blast hole data is available, the improvement may not be as significant since the estimates rely on the conditioning data and less on the technique use to perform the estimation.

Soft boundaries are found in several types of deposits due to the transitional nature of the geological mechanisms involved in the formation of a deposit. There is often some degree of overlap between geological grade controls. Nevertheless conventional grade estimation usually treats the boundaries between geological units as hard boundaries. This is primarily due to the limitations of current estimation and simulation procedures.

Estimation with hard boundaries is straightforward since only the samples within the domain are used. Soft boundaries allow grades from multiple domains to be used in the estimation of each domain. Common practice is to share samples within a given zone of influence of one domain over the other. Samples from different domains are treated as equal to those within the domain; that is, the same mean, variance and covariance model from the samples within the domain are assumed. This generally has the effect of changing the representative statistics of the domain of interest. This corruption of the final grades, especially in the transition zones, often dissuades practitioners from using soft boundaries.

Proposed Methodology

This work presents two different methodologies for grade estimation in the presence of soft boundaries. The first methodology is applicable when the correlation of the variable of interest between two adjacent rock types remains constant within both units and is due to an underlying common factor. One example could be the supergene zone in a porphyry copper deposit (Figure 1.2); supergene enrichment of these systems begin as the portions above the water table are oxidized, transported in solution and precipitated below the ground water table by replacement of pre-existing iron sulfides. The mean and variance in the supergene zone is likely to be higher than the primary zone immediately below, but the spatial correlation structure will remain an underlying common factor because of the original mineralisation.

In this first case of a global soft boundary, it is propose to use a conventional linear model of coregionalization (LMC) to simulate grades using data from adjacent domains. Although the LMC is traditionally used to characterize the spatial variability of multiple properties or metal grades in one domain, we will show that it can be applied to model the spatial variability of one property across the boundary between multiple domains (Figure 1.3). A full model of coregionalization allows us to capture the spatial correlation of grades across the boundaries through a legitimate spatial model that can later be used to cokrige or cosimulate grades using data from adjacent domains. This approach guarantees

the correct reproduction of representative statistics of each geological domain and improves the resource estimation by reducing the uncertainty in transitional zones near boundaries.

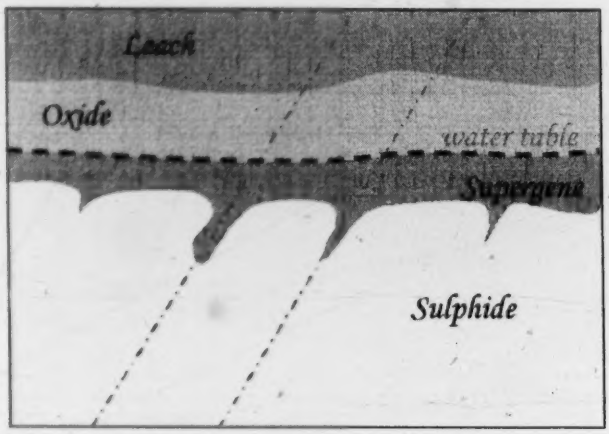

Figure 1.2: Schematic cross-section illustrating porphyry copper mineralisation. The supergene zone develops below the water table as fluids percolate from the oxide zone and downward along faults.

A global stationary soft boundary is often not very realistic. In general soft boundaries are characterized by a non-stationary behavior of the variable of interest near the contact between units; that is, the mean, variance or covariance are not constant within a zone of influence of one rock type into the other and their values depend on the location relative to the boundary (Figure 1.4). There could be an increase or decrease in the mean or variance towards the boundary. Figure 1.5 shows the difference between a global stationary soft boundary and a local non-stationary one. Both domains have a common underlying isotropic covariance model with a range of 100 meters. There may be an increased frequency of fractures towards the boundary between geological domains (Figure 1.6). Faults or brittle zones are examples of this transition. The fractures may be mineralized so the average metal grade will increase closer to the boundary. Alternatively, fractures near the surface of the deposit may be leached by meteoric fluids, which may translate to a decrease in the average grade. The increase in the presence of factures will often lead to an increase or decrease in the variance closer to the boundary.

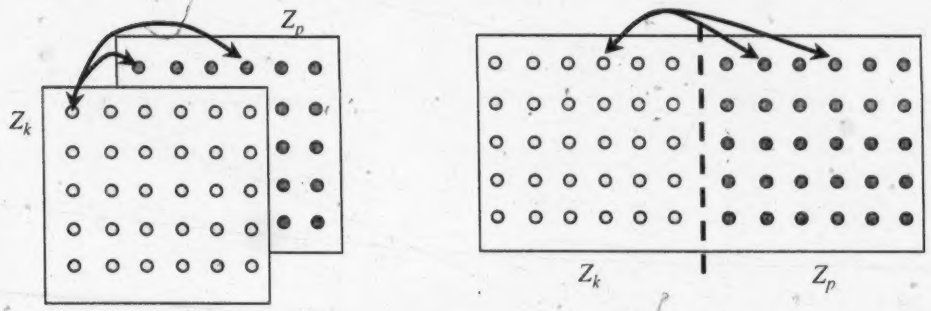

Figure 1.3: Conventional application of a local model of coregionalization (A): model spatial variability of multiple properties, represented by random variables Z_k and Z_p, in one domain. The proposed application (B): model the spatial variability of a single property across a boundary between two geological domains.

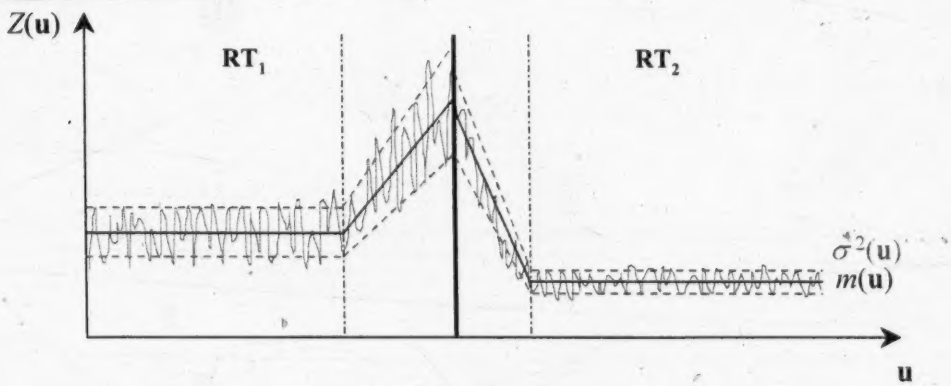

Figure 1.4: Local non-stationary soft boundary between rock type *1* and *2*. The mean and variance are functions of the locations relative to the boundary within a zone of influence of one rock type into the other.

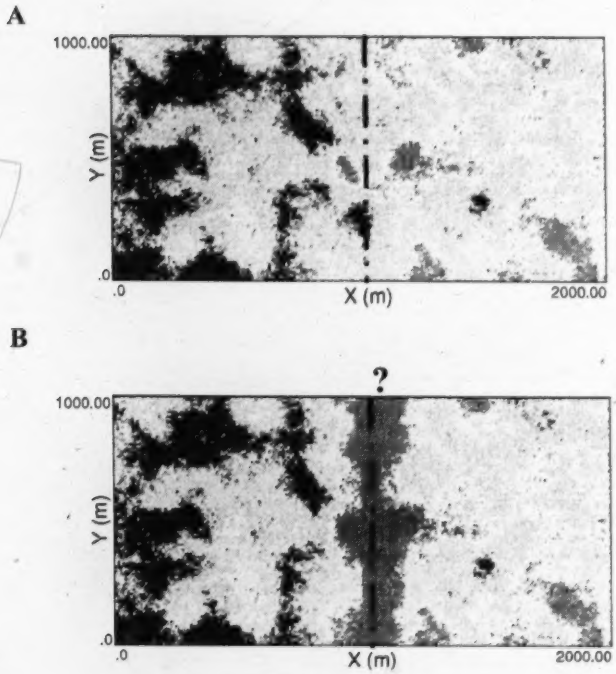

Figure 1.5: Difference between a global soft boundary (A) and a local non-stationary boundary (B). Both domains have a common underlying spatial isotropic structure of 100 meters range.

For local non-stationary soft boundaries it is propose a new technique that accounts for stationary variables within domains and additional non-stationary components near boundaries. The technique involves the following: (i) identification of the rock type and boundary zones based on geological modeling and timing of the different events, (ii) optimization of the non-stationary components of mean and variance in the boundary zone given the stationary statistical parameters of each domain and the data in the boundary region, (iii) decomposition of the covariance model into stationary and non-stationary components of a linear model of coregionalization and optimization of the latest, and (iv) estimation of grades using a non-stationary form of simple cokriging. This technique provides an appealing alternative to capture grade variability for deposits where complex contacts between different rock types exist. The methodology is developed in the context of mining geostatistics, but it is widely applicable in many different settings.

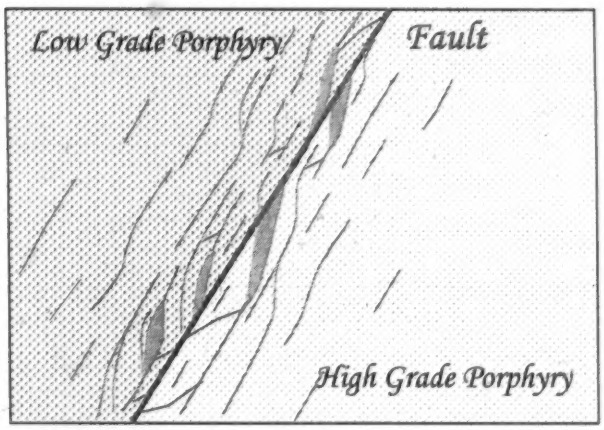

Figure 1.6: Schematic cross-section of a structural boundary between two porphyries. The frequency of fractures increases toward the boundary defined by the fault. Fractures produce a local increase in the mean grade and variability at the boundary.

Thesis Outline

Chapter 2 discusses the application of a conventional linear model of coregionalization to estimate and simulate grades using data from multiple domains. The theoretical background, some small illustrative examples and a full application and comparison with current alternatives are included.

Chapter 3 presents concepts and theoretical background of the proposed technique for estimation in the presence of local non-stationary soft boundaries.

Chapter 4 presents the implementation of the programs involved in the optimization of the non-stationary mean, variance and covariance parameters for cokriging. The structure of these programs is illustrated with a simple 1-D example. A large 3-D example demonstrates the practical application of this technique in Chapter 5.

Finally, discussion on the results of the proposed methodologies, as well as future work and conclusions are presented in Chapter 6.

Chapter 2

Global Linear Model of Coregionalization

As mentioned previously, most geological mechanisms are transitional in nature; therefore, some degree of overlap between geological units can be expected. This translates to correlation of the variable of interest across the geological boundaries.

In this chapter we review how a linear model of coregionalization can be used to model the grade distribution using data from different domains. The theoretical background is illustrated with examples and sensitivity analysis. This proposed methodology is applied to a synthetic deposit and compared to the conventional approach of modeling using hard boundaries.

This methodology assumes that the variable is stationary in each domain. A **global** spatial relationship is used for the grades in different domains. Local non-stationary behavior at the boundary is discussed in the next chapter.

Theoretical Background

The linear model of coregionalization (LMC) provides a method to model the cross covariance of two or more variables. The LMC model is legitimate; that is, the variance of any possible linear combination of these variables is always non-negative. Given a set of K second order stationary random variables, $\{Z_k,\ k=1,...,K\}$, the LMC provides a means to model the cross covariance functions, $Cov_{Z_k Z_p}(\mathbf{h})$, k=1,…,K, p=1,…,K.

Usually Z_k and Z_p represent different properties measured at the same location, for example, gold and copper grades. We consider that each random variable Z_k, $k=1,..,K$ corresponds to the variable of interest in each of the K geological domains or rock types. The model could also be used for multiple grades within multiple rock types.

A linear model of coregionalization assumes that each variable Z_k is a linear combination of n unknown second order stationary independent random variables Y_i with mean m_i and variance σ_i^2, with $i=1,...n$. These n random variables are independent, that is, their cross covariances are zero: $Cov_{Y_i Y_j}(\mathbf{h}) = 0, \forall \mathbf{h}$ and $i \neq j$.

Each Z_k variable is assumed to be a linear sum of the independent factors:

$$Z_k = \sum_{i=1}^{n} a_{ki} Y_i \quad k = 1,...,K \tag{2.1}$$

The coefficients a_{ki} can be positive, negative or zero.

The mean of the k^{th} stationary variable Z_k is:

$$m_k = E\{Z_k\} = E\left\{\sum_{i=1}^{n} a_{ki} Y_i\right\}$$

$$= \sum_{i=1}^{n} a_{ki} m_i$$

The variance can be derived as follows,

$$\sigma_k^2 = E\left\{\left(\sum_{i=1}^{n} a_{ki} Y_i\right)^2\right\} - \left(\sum_{i=1}^{n} a_{ki} E\{Y_i\}\right)^2$$

$$= \left(\sum_{i=1}^{n} a_{ki}^2 E\{Y_i^2\} + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ki} a_{kj} E\{Y_i \cdot Y_j\}\right)$$

$$- \left(\sum_{i=1}^{n} a_{ki}^2 \left(E\{Y_i\}\right)^2 + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ki} a_{kj} E\{Y_i\} \cdot E\{Y_j\}\right)$$

But $E\{Y_i \cdot Y_j\} = E\{Y_i\} \cdot E\{Y_j\}$ since Y_i is independent of Y_j. Then,

$$\sigma_k^2 = \sum_{i=1}^{n} a_{ki}^2 E\{Y_i^2\} + \sum_{i=1}^{n}\sum_{j\neq i}^{n} a_{ki}a_{kj} E\{Y_i\}\cdot E\{Y_j\} - \sum_{i=1}^{n} a_{ki}^2 \left(E\{Y_i\}\right)^2$$
$$-\sum_{i=1}^{n}\sum_{j\neq i}^{n} a_{ki}a_{kj} E\{Y_i\}\cdot E\{Y_j\}$$
$$= \sum_{i=1}^{n} a_{ki}^2 E\{Y_i^2\} - \sum_{i=1}^{n} a_{ki}^2 \left(E\{Y_i\}\right)^2$$
$$= \sum_{i=1}^{n} a_{ki}^2 \sigma_i^2$$

The covariance of Z_k at a vectorial distance **h**, can also be calculated as an expression of the coefficients a_{ki} and the covariances of Y_i for $i=1,...n$:

$$Cov_{Z_k}(\mathbf{h}) = \sum_{i=1}^{n} a_{ki}^2 Cov_{Y_i}(\mathbf{h})$$

Similar to the derivation of the variance, the cross-covariance of Z_k and Z_p, $\forall k\neq p$, with $k,p=1,...,K$ can also be derived as a linear combination of Y_i covariances and coefficients a_{ki} for Z_k, and a_{pj} for Z_p, $i,j=1,...n$,

$$Cov_{Z_kZ_p}(\mathbf{h}) = E\left\{\sum_{i=1}^{n} a_{ki}Y_i(\mathbf{u})\sum_{j=1}^{n} a_{pj}Y_j(\mathbf{u}+\mathbf{h})\right\} - E\left\{\sum_{i=1}^{n} a_{ki}Y_i(\mathbf{u})\right\}\cdot E\left\{\sum_{j=1}^{n} a_{pj}Y_j(\mathbf{u}+\mathbf{h})\right\}$$
$$= \left(\sum_{i=1}^{n} a_{ki}a_{pi}E\{Y_i(\mathbf{u})\cdot Y_i(\mathbf{u}+\mathbf{h})\} + \sum_{i=1}^{n}\sum_{j\neq i}^{n} a_{ki}a_{pj}E\{Y_i(\mathbf{u})\cdot Y_j(\mathbf{u}+\mathbf{h})\}\right)$$
$$-\left(\sum_{i=1}^{n} a_{ki}a_{pi}E\{Y_i(\mathbf{u})\}\cdot E\{Y_i(\mathbf{u}+\mathbf{h})\} + \sum_{i=1}^{n}\sum_{j\neq i}^{n} a_{ki}a_{pj}E\{Y_i(\mathbf{u})\}\cdot E\{Y_j(\mathbf{u}+\mathbf{h})\}\right)$$

But since Y_i is independent of Y_j there is no cross spatial correlation between $Y_i(\mathbf{u})$ and $Y_j(\mathbf{u}+\mathbf{h})$, that is, $E\{Y_i(\mathbf{u})\cdot Y_j(\mathbf{u}+\mathbf{h})\} = E\{Y_i(\mathbf{u})\}\cdot E\{Y_j(\mathbf{u}+\mathbf{h})\}$, $\forall i\neq j$. Then,

$$Cov_{Z_kZ_p}(\mathbf{h}) = \sum_{i=1}^{n} a_{ki}a_{pi}E\{Y_i(\mathbf{u})\cdot Y_i(\mathbf{u}+\mathbf{h})\} - \sum_{i=1}^{n} a_{ki}a_{pi}E\{Y_i(\mathbf{u})\}\cdot E\{Y_i(\mathbf{u}+\mathbf{h})\}$$
$$= \sum_{i=1}^{n} a_{ki}a_{pi}Cov_{Y_i}(\mathbf{h})$$

If the random variables Y_i, $i=1,...,n$ are grouped by those that have the same direct covariance, that is, $C_{Y_i}(\mathbf{h}) = C_{Y_j}(\mathbf{h})$, while the cross terms remain zero due to independence assumption for the Y random function (Journel and Huijbregts, 1978; Goovaerts, 1997), Equation 2.1 can be rewritten as,

$$Z_k = \sum_{i=1}^{n_l} \sum_{l=0}^{L} a_i^l Y_i^l \quad k = 1,...,K$$

where $L+1$ is the number of groups with different direct covariances and n_l is the number of random variables with the same covariance in group l.

The cross covariance of Z_k and Z_p, can also be rewritten as,

$$Cov_{Z_k Z_p}(\mathbf{h}) = \sum_{i=1}^{n_l} \sum_{l=0}^{L} a_{ki}^l a_{pi}^l C^l(\mathbf{h})$$

where the term, $\sum_{i=1}^{n_l} a_{ki}^l a_{pi}^l = b_{kp}^l$ corresponds to the sill contribution of the l^{th} covariance structure, $C^l(\mathbf{h})$.

In summary, a model will be a linear model of coregionalization if all direct and cross-covariances are derived as the linear combinations of l direct covariances $C_{Y_i}(\mathbf{h})$, that is,

$$Cov_{Z_k Z_p}(\mathbf{h}) = \sum_{l=0}^{L} b_{kp}^l C^l(\mathbf{h}) \quad \text{with } b_{kp}^l = b_{pk}^l, l = 1,...,L$$

The coefficient matrix $\left[b_{kp}^l\right]$ must be positive definite. Since this matrix is symmetric, it will be positive definite, if all determinant and sub-determinants, as well as the diagonal terms are non negative (Journel and Huijbregts, 1978).

For a two variable example and replacing the covariances $Cov_{Z_k Z_p}(\mathbf{h})$ and $Cov_{Y_i}(\mathbf{h})$ by the corresponding semi-variograms, the linear model of coregionalization takes the following form:

$$\gamma_{Z_k Z_k}(\mathbf{h}) = \sum_{l=0}^{L} b_{kk}^{l} \cdot \Gamma^{l}(\mathbf{h})$$

$$\gamma_{Z_k Z_p}(\mathbf{h}) = \sum_{l=0}^{L} b_{kp}^{l} \cdot \Gamma^{l}(\mathbf{h})$$

$$\gamma_{Z_p Z_p}(\mathbf{h}) = \sum_{l=0}^{L} b_{pp}^{l} \cdot \Gamma^{l}(\mathbf{h})$$

where $\Gamma^{l}(\mathbf{h})$, $l=1,...,L$ are the nested structures made up of legitimate variogram models and the coefficients b follow the following constrains to ensure a positive definite model:

$$\begin{aligned} b_{kk}^{l} &> 0 \\ b_{pp}^{l} &> 0 \qquad l = 0,...,L \\ b_{kk}^{l} \cdot b_{pp}^{l} &\geq b_{kp}^{l} \cdot b_{kp}^{l} \end{aligned}$$

A linear model of coregionalization, with Z_k and Z_p representing the distribution of the variable of interest in rock type k and rock type p, respectively, is a legitimate spatial correlation model that yields the correct statistics at unsampled locations near the boundary where samples from both domains are used for the estimation or simulation. The calculated LMC spatial model can be used in cokriging or cosimulation to model locations near geological boundaries using samples from adjacent domains. This is a more consistent alternative to the estimation of domains with soft boundaries than assuming the grades are independent or from the same domain.

Illustration of Theory

To illustrate how a linear model of coregionalization can be used to characterize the spatial variability of multiple rock types, consider a 2D example with two domains. The corresponding random variables Z_1 and Z_2 were constructed as a linear combination of three underlying non-standard normal random variables:

$$Z_1 = \sqrt{0.5} \cdot Y_1 + \sqrt{0.5} \cdot Y_2 + 0.0 \cdot Y_3$$

$$Z_2 = \sqrt{0.5} \cdot Y_1 + 0.0 \cdot Y_2 + \sqrt{0.5} \cdot Y_3$$

where,

$$Y_1 \sim \mathrm{N}(0.5,0.5) \text{ with } \gamma_{Y_1}(\mathbf{h}) = 0.05 + 0.45 \cdot Sph_{\left(\begin{smallmatrix} h\max=200 \\ h\min=200 \end{smallmatrix}\right)}(\mathbf{h})$$

$$Y_2 \sim \mathrm{N}(2.0,1.0) \text{ with } \gamma_{Y_2}(\mathbf{h}) = 0.1 + 0.9 \cdot Sph_{\left(\begin{smallmatrix} h\max=50 \\ h\min=300 \end{smallmatrix}\right)}(\mathbf{h})$$

$$Y_3 \sim \mathrm{N}(1.0,0.5) \text{ with } \gamma_{Y_3}(\mathbf{h}) = 0.05 + 0.45 \cdot Exp_{\left(\begin{smallmatrix} h\max=400 \\ h\min=100 \end{smallmatrix}\right)}(\mathbf{h})$$

The random variables Y_i were obtained by unconditional Gaussian simulation for a grid of 1000 by 2000 meters. The simulated values were transformed to a non-standard normal distribution with the corresponding mean and variance values. Ten realizations were simulated.

The covariance of Z_1 and Z_2, obtained from the simulated values, were checked against the analytically derived models:

$$\begin{aligned} Cov_{Z_1}(\mathbf{h}) &= 0.5 \cdot Cov_{Y_1}(\mathbf{h}) + 0.5 \cdot Cov_{Y_2}(\mathbf{h}) \\ &= 0.75 - 0.075 - 0.225 \cdot Sph_{\left(\begin{smallmatrix} h\max=200 \\ h\min=200 \end{smallmatrix}\right)}(\mathbf{h}) - 0.45 \cdot Sph_{\left(\begin{smallmatrix} h\max=50 \\ h\min=300 \end{smallmatrix}\right)}(\mathbf{h}) \end{aligned}$$

$$\begin{aligned} Cov_{Z_2}(\mathbf{h}) &= 0.5 \cdot Cov_{Y_1}(\mathbf{h}) + 0.5 \cdot Cov_{Y_3}(\mathbf{h}) \\ &= 0.5 - 0.05 - 0.225 \cdot Sph_{\left(\begin{smallmatrix} h\max=200 \\ h\min=200 \end{smallmatrix}\right)}(\mathbf{h}) - 0.225 \cdot Exp_{\left(\begin{smallmatrix} h\max=400 \\ h\min=100 \end{smallmatrix}\right)}(\mathbf{h}) \end{aligned}$$

The covariances from the realizations match the analytical models, as it must.

The cross covariance between Z_1 and Z_2 was calculated and checked against its analytical model,

$$\begin{aligned} Cov_{Z_1,Z_2}(\mathbf{h}) &= 0.5 Cov_{Y_1}(\mathbf{h}) \\ &= 0.25 - 0.025 - 0.225 \cdot Sph_{\left(\begin{smallmatrix} h\max=200 \\ h\min=200 \end{smallmatrix}\right)}(\mathbf{h}) \end{aligned} \qquad (2.2)$$

three different spatial arrangement of Z_1 and Z_2 were considered: (1) collocated (just as a check), (2) the two domains adjacent to each other (Figure 2.1A), and (3) the two domains merged (Figure 2.1B) using a categorical binary model obtained via a Boolean simulation program, `ellipsim`, that generate a 2D map of ellipsoids of variable size and anisotropies for a given target proportion (Deutsch and Journel, 1998).

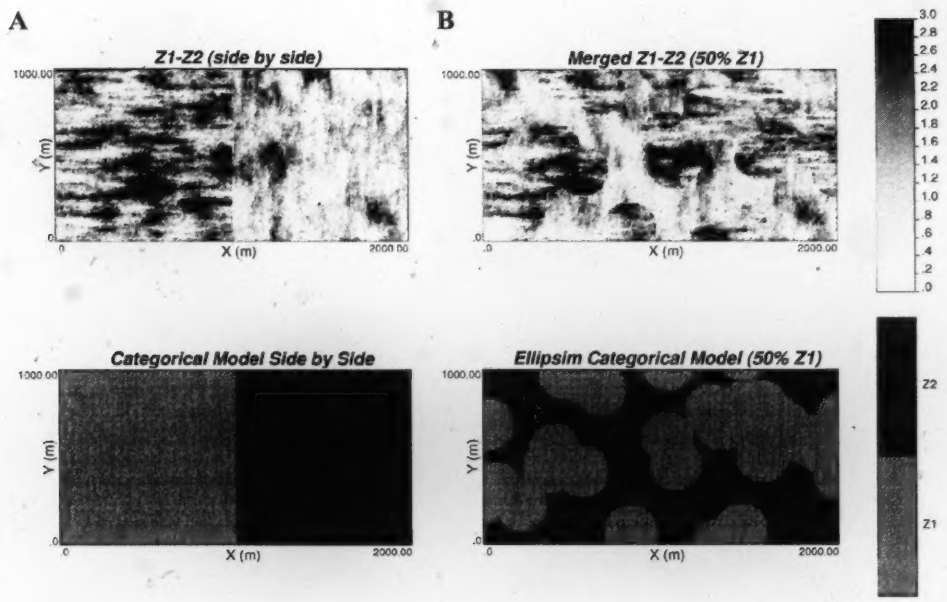

Figure 2.1: Example of two domains and the corresponding categorical models.

As a check of our derivations we compare the cross-covariance between Z_1 and Z_2 when both variables are collocated with the analytical derived model. As shown in Figure 2.2, the average variogram over all realizations is very close to the analytical model (Equation 2.2). The ergodic fluctuations associated with the different realization are relatively small.

In the case where the two domains are side by side, the covariances correspond to the analytical model fairly well (Figure 2.3A), although configurations where the boundary is parallel to the major anisotropy of one of the domains (Z_2 in this case), showed a systematically lower covariance at shorter lag distances than the analytical model (Equation 2.2), and the dispersion of the ergodic fluctuations is greater at lag distances near zero. Inference of the nugget effect of the cross covariance is more uncertain in geometrical configurations similar to this one. However, the nugget effect between the grades at each side of the boundary is not needed for estimation or simulation because data is not collocated nor do we estimate collocated grid blocks.

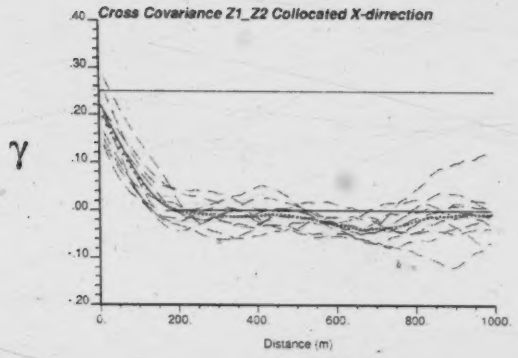

Figure 2.2: Cross-covariance reproduction of the simulated random variables Z_1 and Z_2, assuming both variables are collocated. The dots are the average taken over all realizations; individual realizations are in dashed lines; and the thin red solid line corresponds to the analytical model. The analytical model is very close to the average over all realizations, which makes it difficult to differentiate the dots from the solid line.

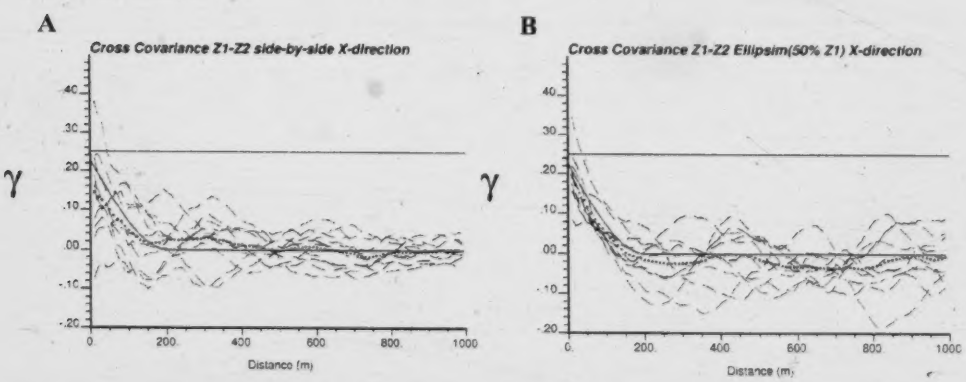

Figure 2.3: (A) Cross-covariance between Z_1 and Z_2 combined side by side. (B) Cross-covariance between Z_1 and Z_2 combined using `ellipsim` categorical model as a boundary model with a target proportion of Z_1 of 50%. The dots are the average taken over all realizations; individual realizations are in dashed lines; and the thin solid line corresponds to the analytical model.

For the second scheme, using a circular shape with radius of 150 meters and three target proportions of 25, 50 and 75%, the cross covariance between the experimental points derived from the average over all realizations compares well with the analytical model (Figure 2.3B) (Equation 2.2). The fluctuations at short lag distances are small. This confirms our expectation that when more contact surfaces between domains are available and are more irregularly oriented; the determination of the nugget effect should have less uncertainty, compared to the

case where a single contact surface exists between the two domains. A completely straight or planar boundary gives the least possible surface area to the boundary. This leads to the smallest possible transition zone between rock types and the fewest possible pairs for variogram calculation. This was also confirmed by a poorer reproduction at shorter lag distances, with lower covariances than the analytical model, when the target proportion of the domain Z_1 was lower than 10%. In addition, when the proportion of one domain decreases, the dispersion of the ergodic fluctuations increases (Figure 2.4).

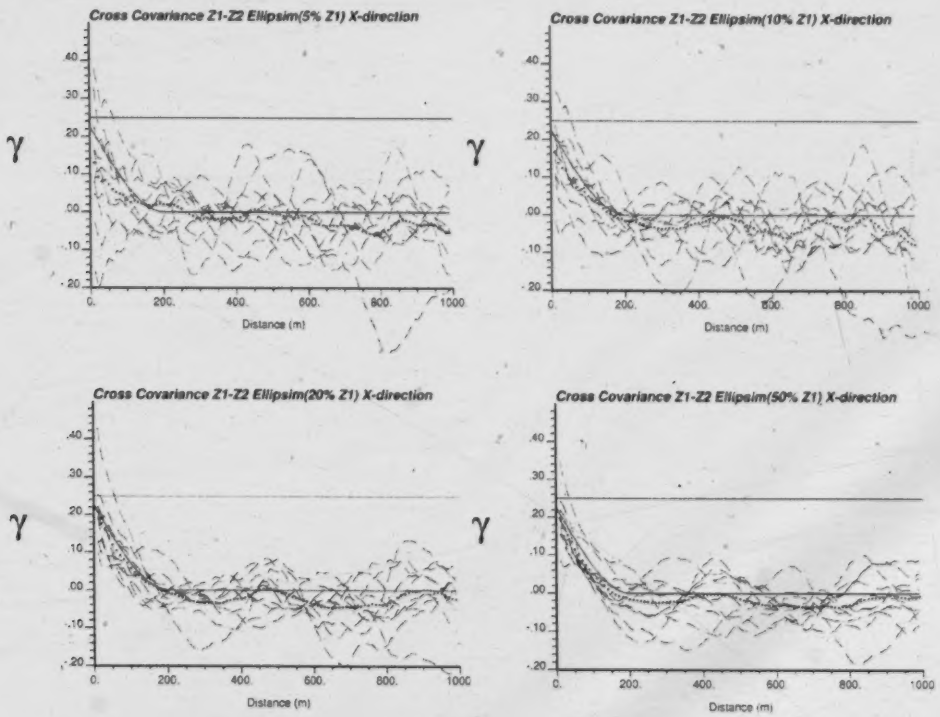
Cross Covariance Z1-Z2 Ellipsim(5% Z1) X-direction
Cross Covariance Z1-Z2 Ellipsim(10% Z1) X-direction
Cross Covariance Z1-Z2 Ellipsim(20% Z1) X-direction
Cross Covariance Z1-Z2 Ellipsim(50% Z1) X-direction
γ
Distance (m)

Figure 2.4: Cross-covariance between Z_1 and Z_2 combined using `ellipsim` categorical model as a boundary model, for target proportion of Z_1 of 5, 10, 20 and 50%. Note that as the target proportion of one of the domains decreases the experimental derived from the average over all showed a systematically lower covariance at shorter lag distances than the dashed analytical model. The dots are the average taken over all realizations; individual realizations are in dashed lines; and the thin solid line corresponds to the analytical model.

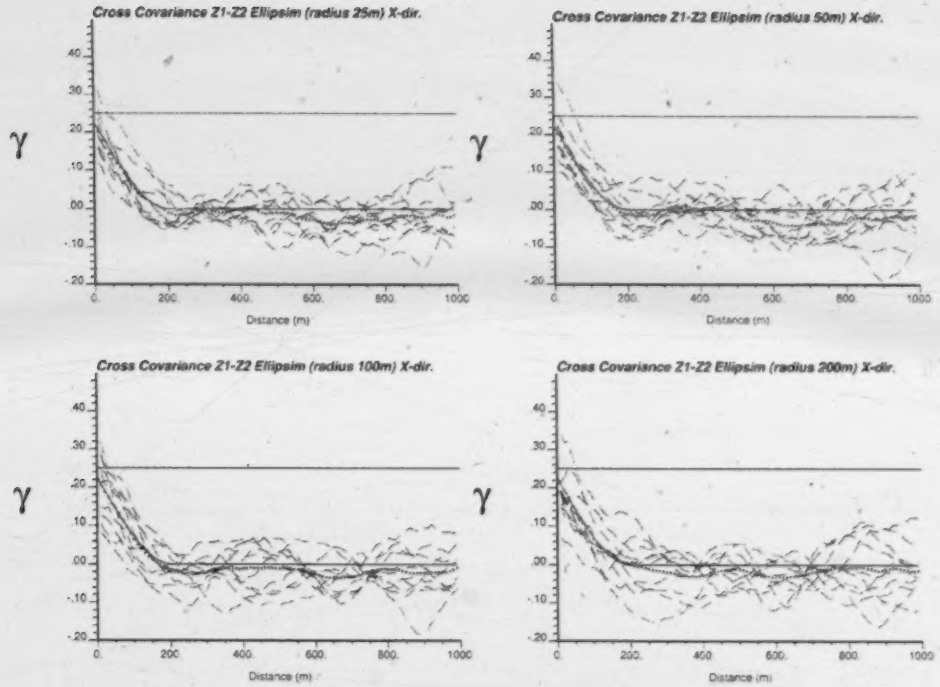
Cross Covariance Z1-Z2 Ellipsim (radius 25m) X-dir.
Cross Covariance Z1-Z2 Ellipsim (radius 50m) X-dir.
Cross Covariance Z1-Z2 Ellipsim (radius 100m) X-dir.
Cross Covariance Z1-Z2 Ellipsim (radius 200m) X-dir.
γ
Distance (m)

Figure 2.5: Cross-covariance between Z_1 and Z_2 combined using `ellipsim` categorical model as a boundary model for different radii: 25, 50, 100 and 200 m. The dots are the average taken over all realizations; individual realizations are in dashed lines; and the thin solid line corresponds to the analytical model.

To determine the influence of the size of the domain, six different radii were checked (25, 50, 100, 150, 200 and 300 meters) using a target proportion of 50%. For every radii, the average cross-covariance between Z_1 and Z_2 is very close to the analytical model and no significant differences were found (Figure 2.5).

Using the same synthetic examples, the impact of different drill hole data spacing was examined. Overall, the reproduction of the cross covariance analytical model is as good as when all simulated values were used, although a wider range of fluctuation between realizations is observed. If the data spacing is larger than the range of the cross-covariance, the calculation of a cross-covariance will be meaningless.

2D Example

A real categorical geological model (Figure 2.6) was used to forecast the results for geometries of a real deposit. The grades, assumed to be percentage of copper, within the five rock types, Z_1 to Z_5 were constructed as a linear combination of four underlying non-standard normal random variables,

$$Y_1 \sim \mathrm{N}(0.01,0.5) \qquad \text{with } \gamma_{Y_1}(\mathbf{h}) = 0.05 + 0.45 \cdot Sph_{\left(\begin{smallmatrix} h\max=100 \\ h\min=100 \end{smallmatrix}\right)}(\mathbf{h})$$

$$Y_2 \sim \mathrm{N}(2.0,1.0) \qquad \text{with } \gamma_{Y_2}(\mathbf{h}) = 0.1 + 0.9 \cdot Sph_{\left(\begin{smallmatrix} h\max=400 \\ h\min=50 \end{smallmatrix}\right)}(\mathbf{h})$$

$$Y_3 \sim \mathrm{N}(0.2,0.75) \qquad \text{with } \gamma_{Y_3}(\mathbf{h}) = 0.05 + 0.70 \cdot Exp_{\left(\begin{smallmatrix} h\max=50 \\ h\min=300 \end{smallmatrix}\right)}(\mathbf{h})$$

$$Y_4 \sim \mathrm{N}(0.75,1.5) \qquad \text{with } \gamma_{Y_4}(\mathbf{h}) = 0.3 + 1.2 \cdot Exp_{\left(\begin{smallmatrix} h\max=400 \\ h\min=250 \end{smallmatrix}\right)}(\mathbf{h})$$

where Y_2 and Y_4 have a 55° anisotropy. The coefficients that multiplied the underlying variables in the summation that originates Z_1 to Z_5 are:

	Z_1	Z_2	Z_3	Z_4	Z_5
Y_1	$\sqrt{0.5}$	0.0	0.0	$\sqrt{0.2}$	$\sqrt{0.7}$
Y_2	0.0	$\sqrt{0.7}$	$\sqrt{0.2}$	0.0	0.0
Y_3	$\sqrt{0.5}$	$\sqrt{0.3}$	0.0	0.0	$\sqrt{0.25}$
Y_4	0.0	0.0	$\sqrt{0.8}$	$\sqrt{0.8}$	$\sqrt{0.05}$

The variables Z_1 up to Z_5 were merged together using the categorical rock type model (Figure 2.7).

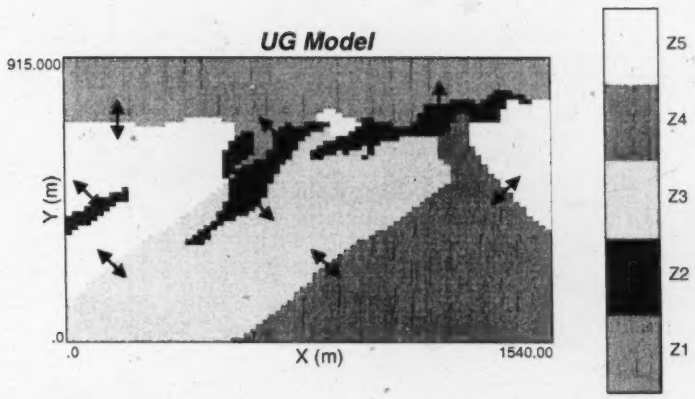

Figure 2.6: Categorical rock type model. The arrows indicate the directions in which the cross-covariance between domains was calculated.

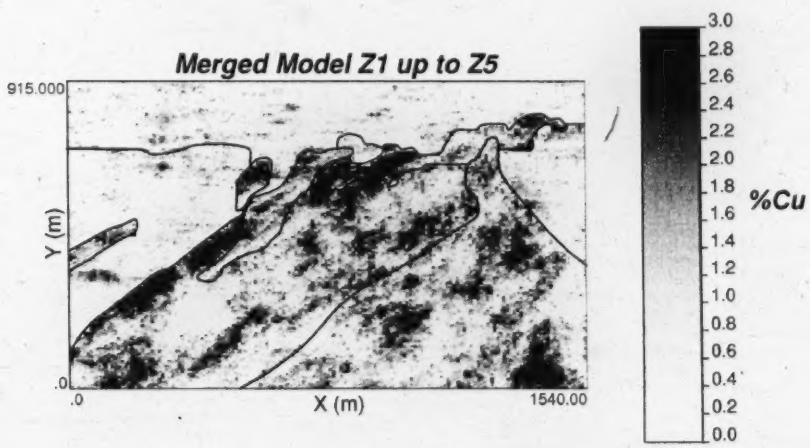

Figure 2.7: A realization of the merged grades Z_1 up to Z_5 using the categorical rock type model of Figure 2.6.

The cross covariance was calculated for each pair of Z_i and Z_j, $i \neq j$, in the directions sketch in Figure 2.6 and compared with the corresponding analytical model:

$$Cov_{Z_1,Z_2}(\mathbf{h}) = 0.29 - 0.019 - 0.271 \cdot Exp_{\left(\begin{smallmatrix} h\max=50 \\ h\min=300 \end{smallmatrix}\right)}(\mathbf{h})$$

$$Cov_{Z_1,Z_3}(\mathbf{h}) = 0.561 - 0.047 - 0.266 \cdot Sph_{\left(\begin{smallmatrix} h\max=100 \\ h\min=100 \end{smallmatrix}\right)}(\mathbf{h}) - 0.248 \cdot Exp_{\left(\begin{smallmatrix} h\max=50 \\ h\min=300 \end{smallmatrix}\right)}(\mathbf{h})$$

$$Cov_{Z_2,Z_3}(\mathbf{h}) = 0.374 - 0.037 - 0.337 \cdot Sph_{\left(\begin{smallmatrix} h\max=400 \\ h\min=50 \end{smallmatrix}\right)}(\mathbf{h})$$

$$Cov_{Z_2,Z_5}(\mathbf{h}) = 0.206 - 0.014 - 0.192 \cdot Exp_{\left(\begin{smallmatrix} h\max=50 \\ h\min=300 \end{smallmatrix}\right)}(\mathbf{h})$$

$$Cov_{Z_3,Z_4}(\mathbf{h}) = 1.2 - 0.24 - 0.96 \cdot Exp_{\left(\begin{smallmatrix} h\max=400 \\ h\min=250 \end{smallmatrix}\right)}(\mathbf{h})$$

$$Cov_{Z_3,Z_5}(\mathbf{h}) = 0.3 - 0.06 - 0.24 \cdot Exp_{\left(\begin{smallmatrix} h\max=400 \\ h\min=250 \end{smallmatrix}\right)}(\mathbf{h})$$

$$Cov_{Z_4,Z_5}(\mathbf{h}) = 0.487 - 0.079 - 0.168 \cdot Sph_{\left(\begin{smallmatrix} h\max=100 \\ h\min=100 \end{smallmatrix}\right)}(\mathbf{h}) - 0.24 \cdot Exp_{\left(\begin{smallmatrix} h\max=400 \\ h\min=250 \end{smallmatrix}\right)}(\mathbf{h})$$

The experimental cross covariance obtained from the average overall realizations compares very well with the analytical models (Figure 2.8), except for the pairs Z_3/Z_4 and Z_4/Z_5, that have a side by side arrangement that shows lower covariances at shorter lag distances.

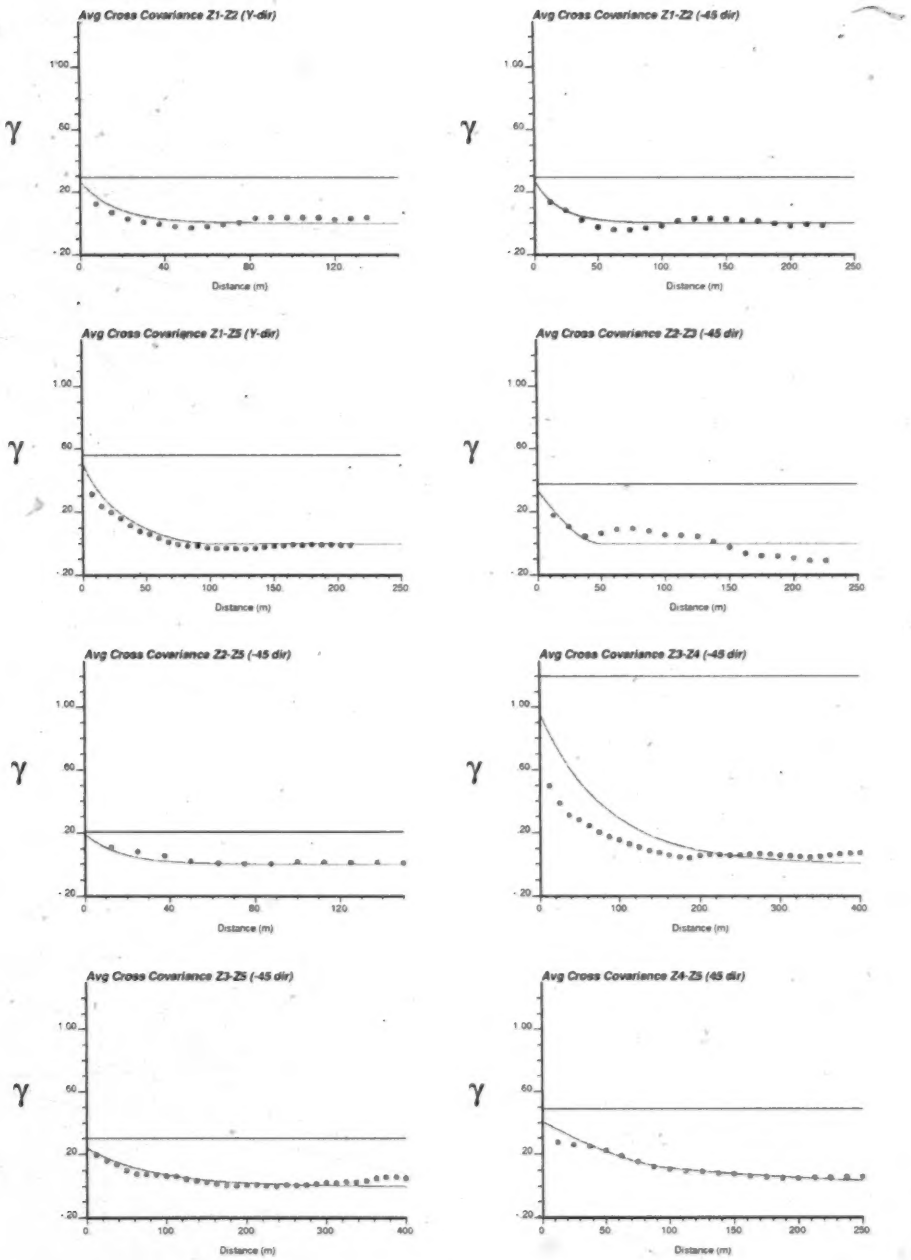
Avg Cross Covariance Z1-Z2 (Y-dir)
Avg Cross Covariance Z1-Z2 (-45 dir)
Avg Cross Covariance Z1-Z5 (Y-dir)
Avg Cross Covariance Z3-Z3 (-45 dir)
Avg Cross Covariance Z2-Z5 (-45 dir)
Avg Cross Covariance Z3-Z4 (-45 dir)
Avg Cross Covariance Z3-Z5 (-45 dir)
Avg Cross Covariance Z4-Z5 (45 dir)
γ
Distance (m)

Figure 2.8: Cross-covariance reproduction of the simulated pairs Z_i and Z_j for $i \neq j$, combined by the categorical rock type model. The experimental points correspond to the average over ten realizations, and the thin solid line corresponds to the analytical model.

Application

A synthetic example was created in order to use a full LMC cosimulation and compare it with the results obtained from simulating two adjacent rock types independently. The LMC model was obtained by calculating the cross variograms between values of the different domains and the direct variograms within each rock type.

Using a similar methodology, the linear combination of three underlying standard normal random variables were used to populate a synthetic geological model; this will be considered as the 'true' image (Figure 2.9) for comparison:

$$Z_1 = \sqrt{0.5} \cdot Y_1 + \sqrt{0.5} \cdot Y_2 + 0.0 \cdot Y_3$$
$$Z_2 = \sqrt{0.5} \cdot Y_1 + 0.0 \cdot Y_2 + \sqrt{0.5} \cdot Y_3$$

with,

$$Y_1 \sim \mathrm{N}(0,1) \text{ with } \gamma_{Y_1}(\mathbf{h}) = 0.1 + 0.9 \cdot Sph_{\left(\begin{smallmatrix} h\max=200 \\ h\min=200 \end{smallmatrix}\right)}(\mathbf{h})$$
$$Y_2 \sim \mathrm{N}(0,1) \text{ with } \gamma_{Y_2}(\mathbf{h}) = 0.1 + 0.9 \cdot Sph_{\left(\begin{smallmatrix} h\max=50 \\ h\min=300 \end{smallmatrix}\right)}(\mathbf{h})$$
$$Y_3 \sim \mathrm{N}(0,1) \text{ with } \gamma_{Y_3}(\mathbf{h}) = 0.1 + 0.9 \cdot Exp_{\left(\begin{smallmatrix} h\max=400 \\ h\min=100 \end{smallmatrix}\right)}(\mathbf{h})$$

Later the random variables Z_1 and Z_2 were transform to lognormal distributions to resemble a more realistic distribution of grades assumed to be copper,%:

$$Z_1 \sim \text{lognormal } (m=1, \sigma=1)$$
$$Z_2 \sim \text{lognormal } (m=2, \sigma=2)$$

The 2D reference image (2000 by 1000 meters, with a 10 meter grid spacing in both directions) was sampled at a spacing of 70 meters in the X-direction and 10 meters in the Y-direction yielding a total of 2800 samples.

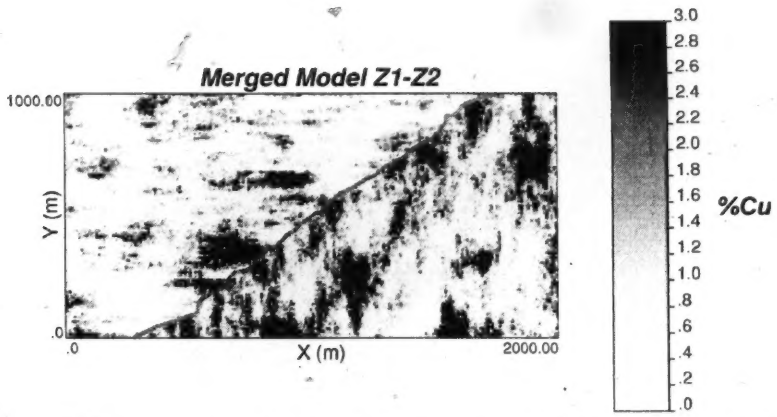

Figure 2.9: Reference or 'true' image. Z_1 is represented by $\mathbf{RT_1}$ (left), Z_2 by $\mathbf{RT_2}$ (right).

Variography

Variograms were calculated from the normal scores transform values from each rock type, $\mathbf{RT_1}$ and $\mathbf{RT_2}$. Cross variograms can not be calculated if the variables are not collocated, which is the case here since we are trying to characterize the spatial variability across the boundary between $\mathbf{RT_1}$ and $\mathbf{RT_2}$. An alternative (Wawruch *et. al.* 2003) is to (1) calculate the cross covariance between the variables, (2) extrapolate the experimental points at lags near to zero to obtain the structured cross covariance ($B_{Z_1-Z_2}$) (Figure 2.10), (3) determine the relative nugget effects for Z_1 and Z_2, and (4) calculate the sill of the cross variogram between Z_1 and Z_2 as:

$$Cov_{Z_1-Z_2}(0) = \frac{B_{Z_1-Z_2}}{1-\left[\frac{1}{2}\left(\frac{Cov^0_{Z_1}}{\sigma^2_{Z_1}}+\frac{Cov^0_{Z_2}}{\sigma^2_{Z_2}}\right)\right]}$$

Note the superscript 0 ($i = 0$) is used to represent the variance contribution due to the nugget effect. Then $Cov^0_{Z_i}/\sigma^2_{Z_i}$ is the relative nugget effect for Z_i.

In this example, the relative nugget effects obtained from the direct variograms of each rock type were both 0.1, the structured cross covariance was chosen at 0.4, so the sill of the cross variogram is 0.44. With this value the experimental points

from the cross covariance can be inverted to obtain the cross variogram between Z_1 and Z_2. Most cokriging and cosimulation programs require the LMC to be defined with variogram models, which requires the nugget effect and the sill of the cross variogram, nevertheless since there are no collocated data nor do we estimate collocated grid blocks, the nugget effect between the grade at each side of the boundary is not needed in any calculations.

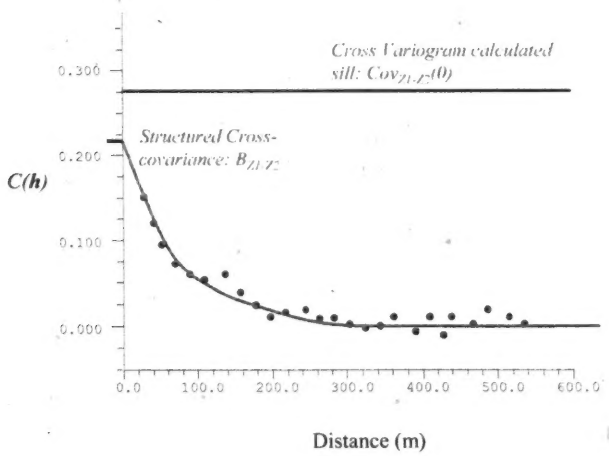

Figure 2.10: Sketch with the structured cross covariance and calculated sill of a cross variogram given an experimental cross covariance between two non-collocated variables.

The direct and cross variograms of Z_1 and Z_2 were modeled using a linear model of coregionalization (Table 2.1 and Figure 2.11) obtained by a semi-automatic variogram fitting program (Larrondo *et. al.*, 2003). Since independent simulations of Z_1 and Z_2 were also performed, the direct variograms of each variable were modeled independently (Table 2.2 and Figure 2.12).

Variable	Structure Type	Sill Contribution	Range Hmax	Range Hmin
Z1	Nugget Effect	0.100		
	Spherical	0.102	125.0	83.0
	Exponential	0.600	125.0	290.0
	Spherical	0.198	433.0	290.0
Z2	Nugget Effect	0.100		
	Spherical	0.524	125.0	83.0
	Exponential	0.005	125.0	290.0
	Spherical	0.371	433.0	290.0
Z1-Z2	Nugget Effect	0.100		
	Spherical	0.231	125.0	83.0
	Exponential	0.053	125.0	290.0
	Spherical	0.056	433.0	290.0

Table 2.1: LMC model for Z_1 and Z_2.

Variable	Structure Type	Sill Contribution	Range Hmax	Range Hmin
Z1	Nugget Effect	0.10		
	Exponential	0.35	51.0	241.0
	Spherical	0.46	175.0	241.0
	Exponential	0.09	2010.0	2340.0
Z2	Nugget Effect	0.10		
	Spherical	0.36	288.0	91.8
	Exponential	0.37	288.0	192.0
	Exponential	0.17	288.0	244.0

Table 2.2: Variogram models for Z_1 and Z_2, obtained considering the two variables independently.

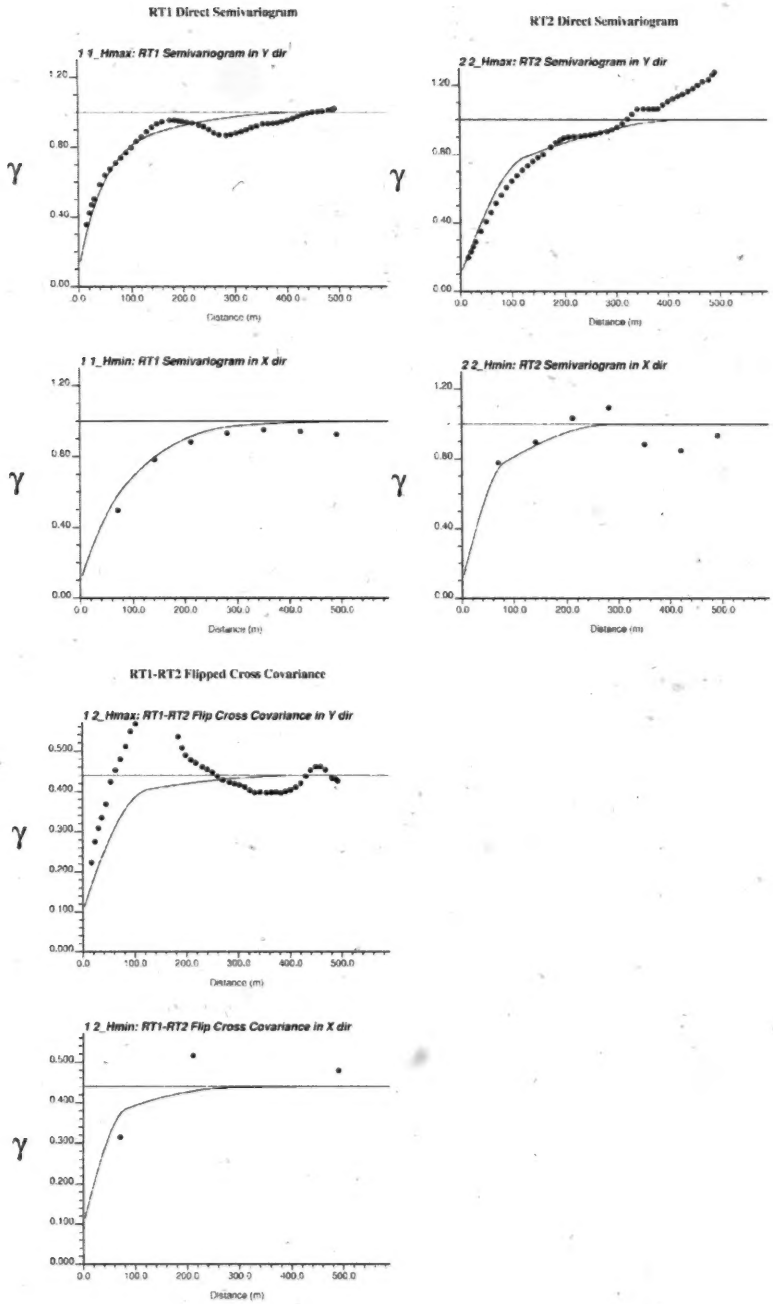
RT1 Direct Semivariogram
1 1_Hmax: RT1 Semivariogram in Y dir
RT2 Direct Semivariogram
2 2_Hmax: RT2 Semivariogram in Y dir
1 1_Hmin: RT1 Semivariogram in X dir
2 2_Hmin: RT2 Semivariogram in X dir
RT1-RT2 Flipped Cross Covariance
1 2_Hmax: RT1-RT2 Flip Cross Covariance in Y dir
1 2_Hmin: RT1-RT2 Flip Cross Covariance in X dir
γ
Distance (m)

Figure 2.11: Direct and cross variograms and LMC model for Z_1 and Z_2.

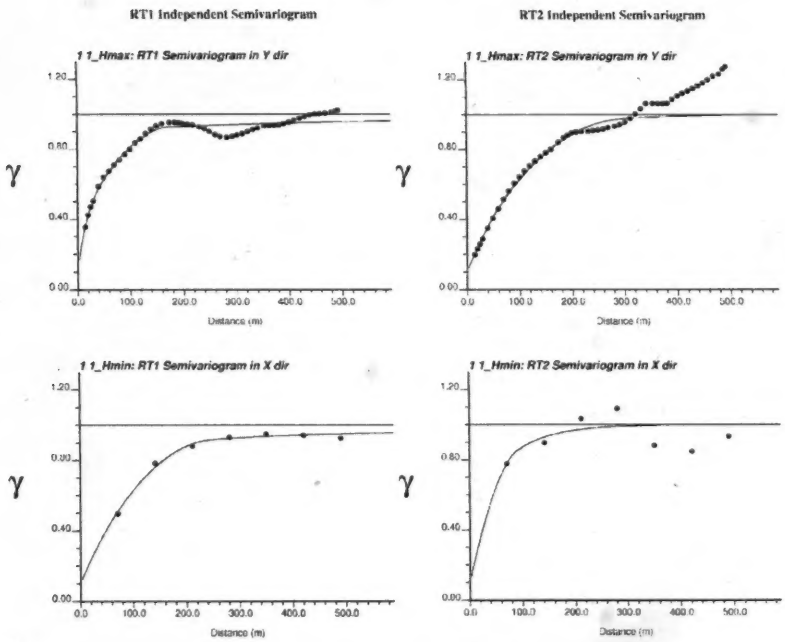
RT1 Independent Semivariogram
RT2 Independent Semivariogram
1 1_Hmax: RT1 Semivariogram in Y dir
1 1_Hmax: RT2 Semivariogram in Y dir
1 1_Hmin: RT1 Semivariogram in X dir
1 1_Hmin: RT2 Semivariogram in X dir
γ
Distance (m)

Figure 2.12: Variogram model for Z_1 and Z_2, obtained independently.

Simulation

The cosimulation was performed using the full LMC cokriging option of the `ultimate sgsim` program (Deutsch and Zanon, 2002) and the parameters in Table 2.3; in this case each rock type was simulated using the samples of the other rock type, as a secondary variable. For the comparative case, sequential Gaussian simulation (Table 2.4) was used to simulate each rock type independently as the contact between $\mathbf{RT_1}$ and $\mathbf{RT_2}$ was treated as a hard boundary. Although for $\mathbf{RT_1}$ the independent variogram model shows a zonal anisotropy in the minor direction, the sill contribution associated with this is only 0.09 so the search radius was chosen at 300 meters.

Number of Nodes	*24, maximum of 6 per octant*
Search Radius	*450 x 450 x 10 m*
Number of Realizations	*10*

Table 2.3: Parameters used for cosimulation of Z_1 and Z_2.

Number of Nodes	*24, maximum of 6 per octant*
Search Radius	*300 x 300 x 10 m*
Number of Realizations	*10*

Table 2.4: Parameters used for independent simulations of Z_1 and Z_2.

Validation

The reproduction of the direct variograms for both the cosimulation and independent simulation was fairly good (Figure 2.13). Although the reproduction of the cross variogram was poor compared with the analytical model, the first 100 meters (total range) in the X-direction showed a similar amount of correlation (Figure 2.14). The case where the contact between $\mathbf{RT_1}$ and $\mathbf{RT_2}$ was assumed to be a hard boundary, resulted in almost no correlation for lags less than the range of the cross variogram, and is significantly lower than the correlation of the conditional data across the boundary. This correlation is a remnant correlation from data, not from modeling. While for a soft boundary assumption, the correlation of the average over all realizations is closer to the correlation shown by the 'truth' reference.

The validation of the parameters and variogram models used for independent simulation and cosimulation was done estimating a location where the value was removed from the dataset. This is done iteratively for all samples, but removing the entire drill hole. The estimation was done using the program `kt3d` in a cross validation mode for independent simulations, and a modified version of `cokb3d` program for cosimulation. Both programs return the estimated values for each sample location, the estimation variance and the error (true minus the estimate). With these results we can check the reproduction of the values, the distribution of the errors and whether the distributions of uncertainty are accurate and precise. Accuracy is achieved when the fraction of true values that fall in a probability interval of width p exceeds p for all p in [0,1] (Deutsch 2002). Precision is how close the fraction of true values is equal to p for all p in [0,1]. These properties are reflected in an accuracy plot (Figure 2.15) by the closeness of the points to the 45° line, if the points fall on the line, the distribution representing the model is accurate and precise, if the points fall above the line the distribution is accurate but not precise, if they fall below the model is neither accurate nor precise.

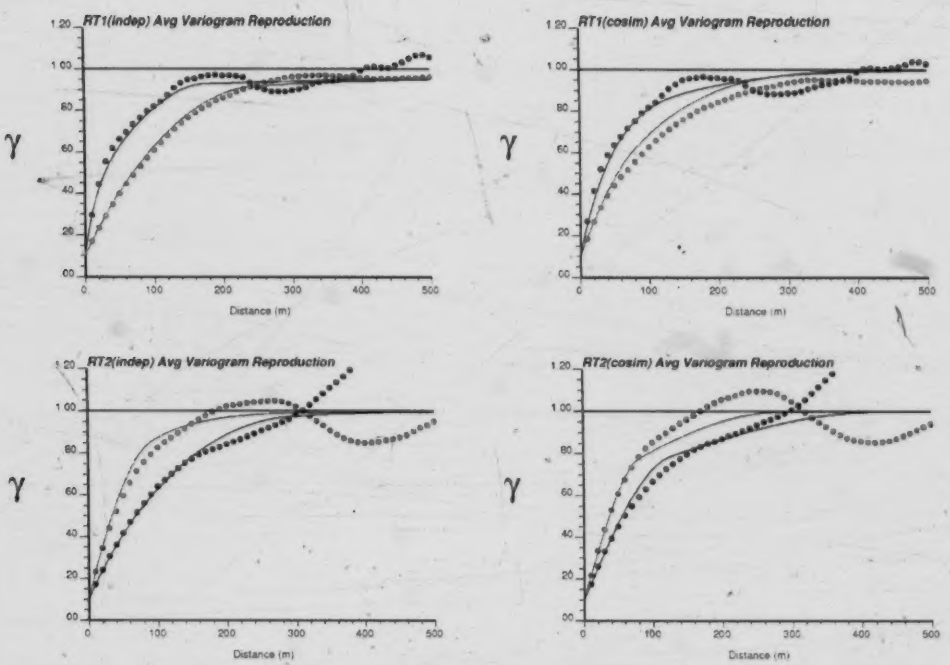

Figure 2.13: Direct variograms reproduction for Z_1 and Z_2, cosimulated (right) and independently simulated (left). The dots represent the average of simulated values over ten realizations and the solid line is the analytical model derived from the theoretical expression.

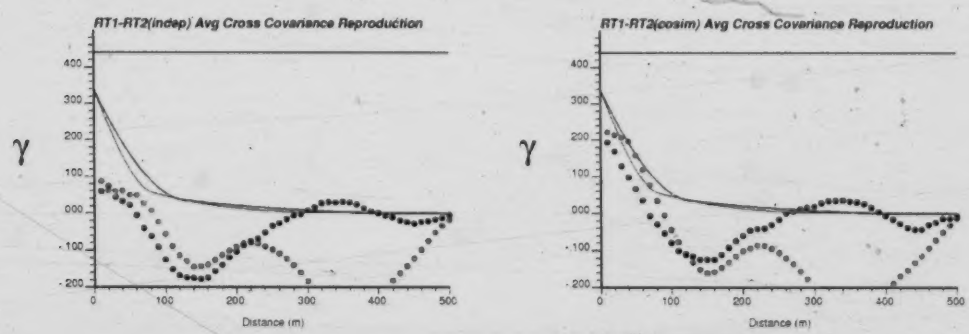

Figure 2.14: Cross covariance reproduction for Z_1 and Z_2, cosimulated (right) and independently simulated (left). In a soft boundary scheme (right) the correlation between the simulated values is very close to the 'truth' reference. In the hard boundary assumption, the correlation at short lag distances is significantly lower. The dots represent the average of simulated values over ten realizations and the solid line is the analytical model derived from the theoretical expression.

Cross validation of the model obtained by independently simulating Z_1 and Z_2 showed that the model is accurate and precise. The cosimulated model is also accurate, and equally precise for $\mathbf{RT_1}$, while for $\mathbf{RT_2}$ is slightly less precise than the model obtained from independent simulations (Figure 2.15). This is not surprising since the fitted LMC model for this rock type did not fit the data as well as for $\mathbf{RT_1}$. This is a common disadvantage when using a linear model of coregionalization. The cosimulated model did, however, show less smoothing (Figure 2.16), which translates to less conditional bias in the estimation.

Although both models are similarly accurate and precise the overall uncertainty, defined as the average kriging variance of all samples (Deutsch 2002), is significantly lower for the cosimulated model (0.3 for both Z_1 and Z_2) than for the independently simulated model (0.62 for Z_1 and 0.91 for Z_2).

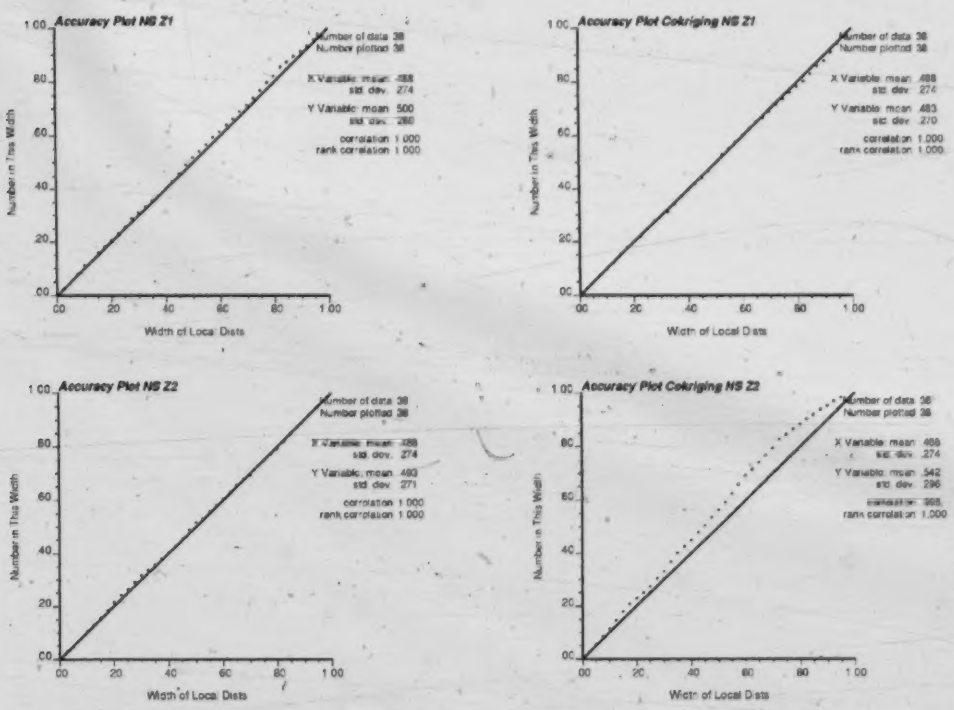

Figure 2.15: Accuracy plot for $Z_1$and Z_2, estimate independently (left) versus cosimulated (right). Cross validation show that the models from independent simulation or cosimulation of Z_1 are accurate and precise. For Z_2 the parameters used for cosimulation results a slightly less precise model than in the case of independent simulation.

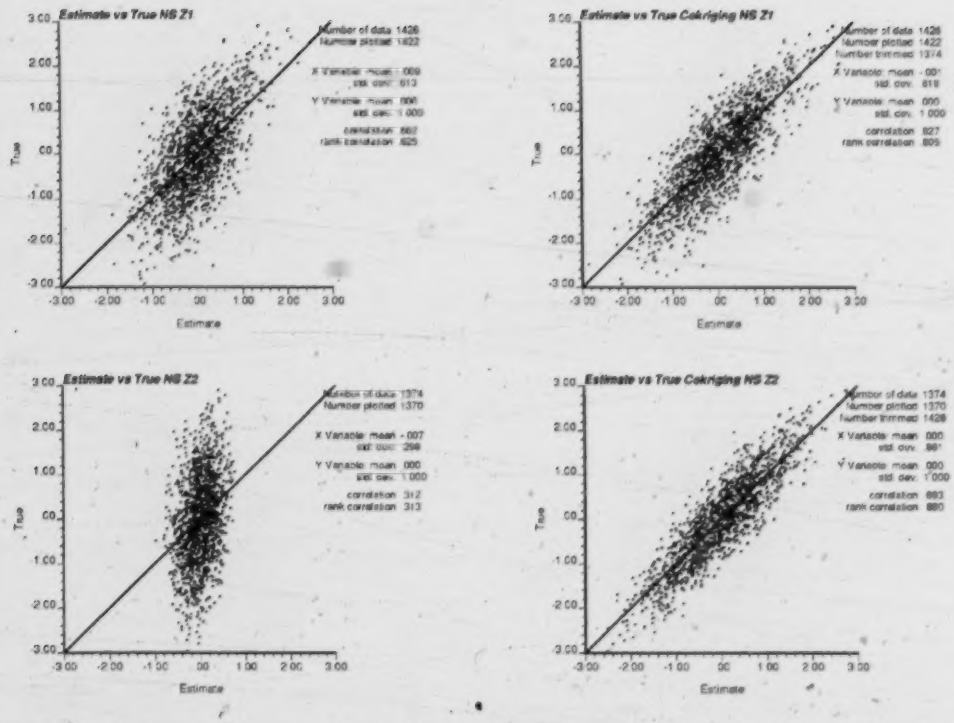
Estimate vs True NS Z1
Estimate vs True Cokriging NS Z1
Estimate vs True NS Z2
Estimate vs True Cokriging NS Z2
True
Estimate

Figure 2.16: Cross validation of data values in $\mathbf{RT_1}$ and $\mathbf{RT_2}$, estimate independently (left) versus cosimulated (right). The cokriging cross validation shows far less conditional bias and a much higher correlation than the estimation of each rock type independently, especially for $\mathbf{RT_2}$.

The distribution of errors (true-estimated) should be symmetric and centered at zero, as occurs for both schemes, but the standard deviation of the errors for cokriged estimates is significantly lower than independently kriged values, as shown in Figure 2.17.

The cumulative distribution of back transformed simulated values shows very good reproduction of the data histograms, for both schemes. The target mean and variance are well reproduced for both cosimulation and independent simulations (Figure 2.18).

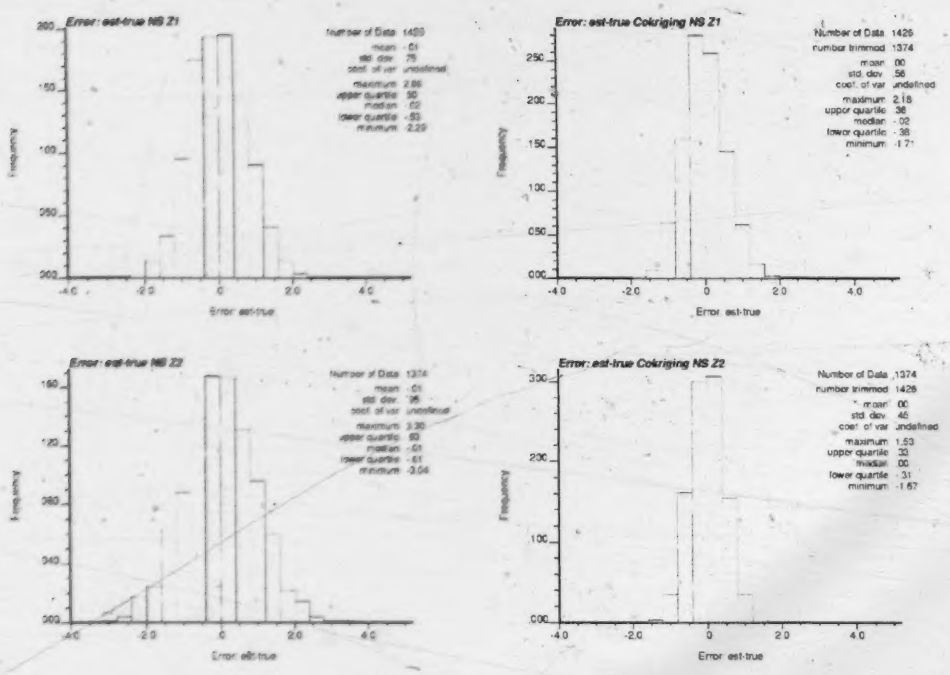

Figure 2.17: Error (true-estimated) distribution for Z_1 and Z_2, estimate independently (left) versus cosimulated (right). The cosimulation scheme shows significantly lower standard deviation of the errors, in both rock types, than independent simulation.

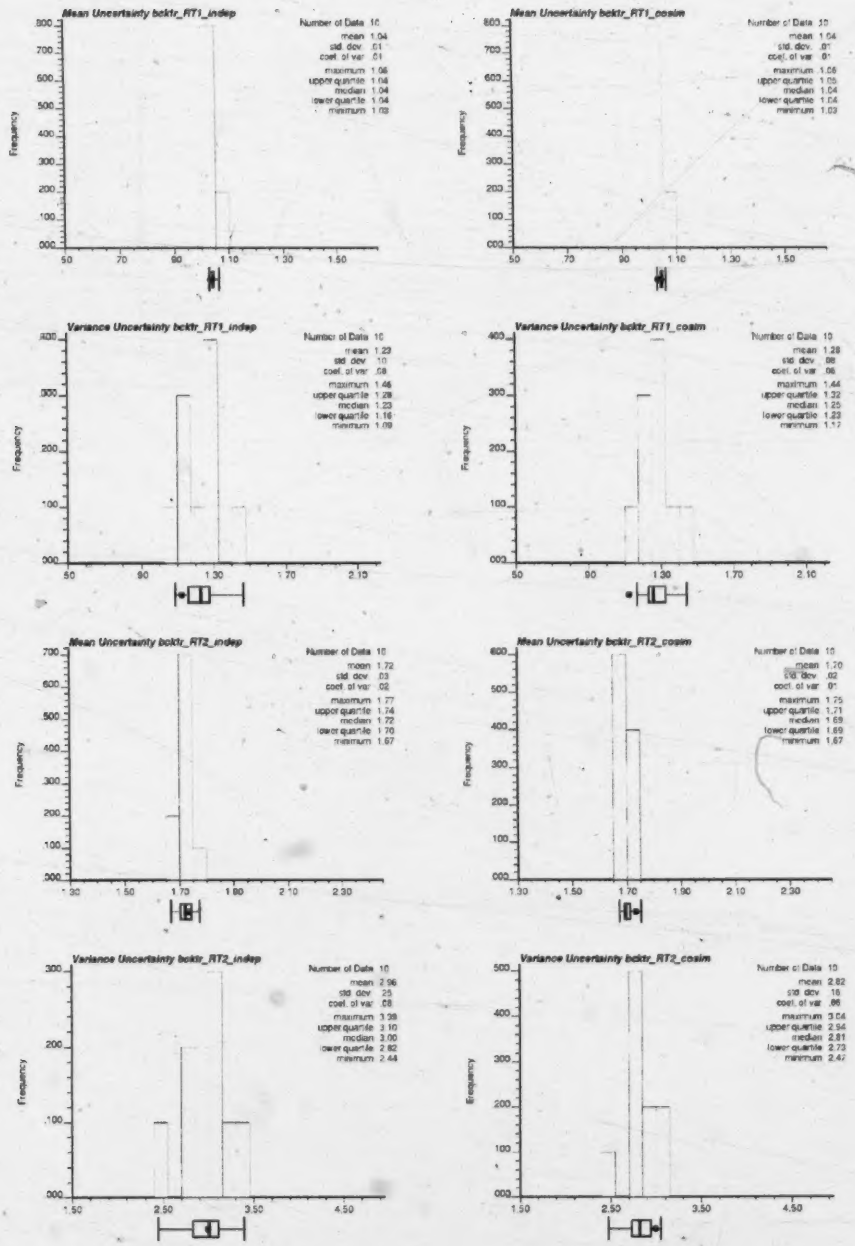

Figure 2.18: Global mean and variance reproduction for Z_1 and Z_2; cosimulation (right side) and independent simulation (left side). The dots represent the target values obtained from the conditioning data.

Comparison at the boundary

In order to compare the performance of the two methods, we need to focus on the results near the boundary where we can expect to have greater differences.

One comparison was done using the expected value (E-type value) in original units at each location compared to the 'true' value in the reference map. The expected value is the average of the simulated realizations at each location. The block values obtained from cosimulation show systematically higher correlation coefficients with the true values. As expected, the difference between the two methods becomes smaller beyond the range of correlation of the cross variogram (Figure 2.19).

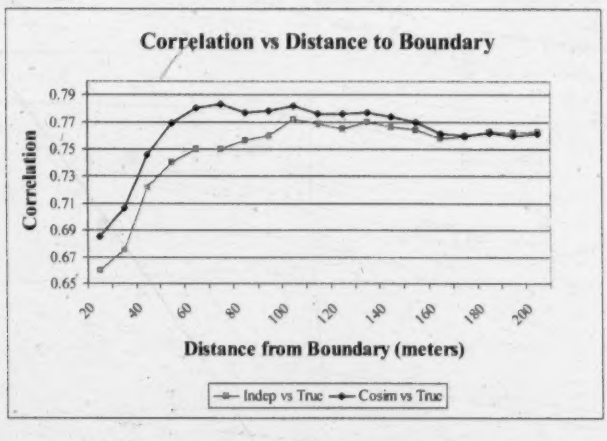

Figure 2.19: Correlation coefficient between E-type estimates of cosimulated and independently simulated models, and the "true" values considering blocks within a given distance from the boundary between Z_1 and Z_2. The higher correlation coefficient with the true values shown by the blocks estimated by cosimulation indicate this model better represents the underlying correlation that exists between Z_1 and Z_2.

The variance of the blocks from each realization within a given distance from the boundary was also compared. The average of the variance over all the realizations showed lower variance for the block values obtained using cosimulation (Figure 2.20). This variance is also closer to the average variance calculated from the same group of blocks in the 'true' reference map.

The methodology proposed in this section has the advantage of improved resource estimation by reducing the uncertainty in transitional zones near boundaries and reproducing the correlation of the conditioning data across a soft

boundary. It also shows a decrease of smoothing in the estimates if kriging is the tool to obtain the resources.

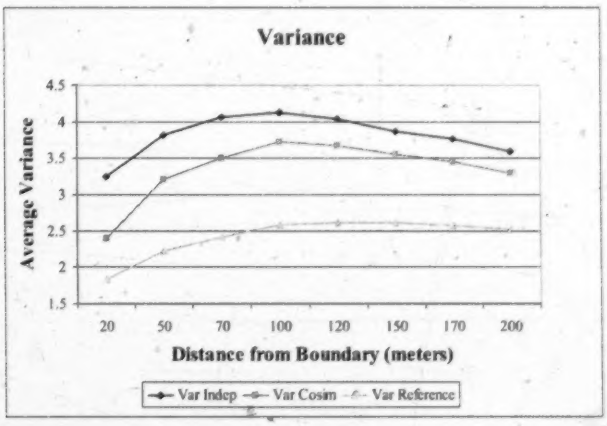

Figure 2.20: Average variance calculated from blocks within a given distance from the boundary between Z_1 and Z_2. The average variance obtained from cosimulation is lower than the variance obtained from independent simulations. The difference between the two methods decreases as we consider blocks further away from the boundary, where the influence of the data from the adjacent rock type in the cosimulation decreases. The average variance from cosimulated blocks is closer to the variance of the same blocks in the 'true' reference map, than the average variance obtained from independent simulations.

Chapter 3

Local Non-stationary Model of Coregionalization

The global LMC described in the previous chapter assumes that the grades are stationary within each domain and any cross correlation is also stationary and extends across the entire domain. This chapter will describe a new methodology to account for non-stationary components of the mean, variance and covariance near the boundaries. It will be show that non-stationary features in the vicinity of a boundary can be parameterized into a local model of coregionalization. With a legitimate spatial model, estimation of grades can be performed using a form of non-stationary cokriging.

Theoretical Background

For the case of a geological model with K rock types or domains, there are a maximum of $K(K-1)/2$ boundary zones to be defined. These regions may be characterized by a non-stationary behavior of the variable of interest. The zones of influence of one region into an adjacent one are not necessarily symmetric on each side of the boundary and there may be cases of truly hard boundaries where there are no special features at the boundary. When more than two rock types converge at a boundary, two or more rock types may influence the boundary zone in the adjacent domain. In this case, precedence or ordering rules should determine the dominant boundary zone. Only one non-stationary factor will be considered dominant at each location.

The extent of the boundary zones and a set of precedence rules are required to partition a geological model into stationary and non-stationary regions. Each of these concepts will be described below.

The continuous random function $Z(\mathbf{u})$ that represents the distribution of the property of interest, e.g. metal content, over the area of study, can be decomposed into K stationary random variables Z_k $k=1,...,K$ and a maximum of $K(K-1)/2$ non-stationary boundary variables $Z_{kp}(\mathbf{u})$, with $k,p=1,...K$ and $Z_{kp}(\mathbf{u})=Z_{pk}(\mathbf{u})$. Then, at all locations, $Z(\mathbf{u})$ can be explained by the sum of a stationary component from the collocated rock type and perhaps a single non-stationary boundary variable, that is,

$$Z(\mathbf{u}) = Z_k + Z_{kp}(\mathbf{u}) \quad \text{where } \mathbf{u} \in \text{ rock type } k \tag{3.1}$$

and p is a surrounding rock type that shares a boundary with rock type k ($\mathbf{RT_k}$).

Maximum Distance of Influence

The maximum distance of influence orthogonal to the boundary of rock type k into rock type p is denoted $dmax_{kp}$. A boundary zone is defined by two distances: $dmax_{kp}$ and $dmax_{pk}$, since there is no requirement that the regions at each side of the boundary are symmetric, that is, $dmax_{kp} \neq dmax_{pk}$. Figure 3.1 shows a schematic illustration of this.

The modeler using all geological information available and its expertise should establish these distances. Creating an average value profile against distance to boundary could be helpful in identifying the zones of influence between domains. An automatic optimization algorithm is unlikely to work given that the stationary portion of the mean and variance, as well as the non-stationary factors that affect these statistics, are also unknown.

All maximum distances can be arranged in a non-symmetric square K by K matrix:

$$\begin{bmatrix} -1 & dmax_{12} & \cdots & dmax_{1K} \\ dmax_{21} & \ddots & & \vdots \\ \vdots & & \ddots & dmax_{K-1\,K} \\ dmax_{K1} & \cdots & dmax_{K\,K-1} & -1 \end{bmatrix}$$

The maximum distance of influence of a rock type to itself is meaningless and any default value could be assigned (-1 in the sketch above).

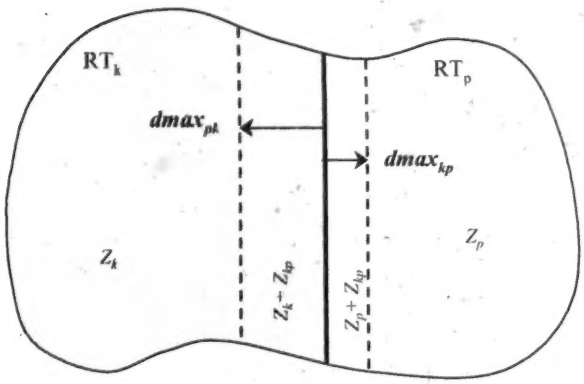

Figure 3.1: Maximum distances of influence for the boundary zone between rock type k ($\mathbf{RT_k}$) and rock type p ($\mathbf{RT_p}$).

A rock type is assigned at all locations within the geological model. Where the distance to the boundary is lower than the maximum distance of influence of the adjacent rock, a boundary zone rock type is also assigned. We will denote the region corresponding to the zone of influence of rock type p into rock type k (Figure 3.2) as the boundary zone $\mathbf{BZ_{pk}}$ and it is defined as all locations $\mathbf{u}_i$ in rock type k, such that $d_{pk}(\mathbf{u}_i) \leq dmax_{pk}$, where $d_{pk}(\mathbf{u})$ is the shortest distance to k-p boundary. Equivalently, the region on the other side of the boundary between rock type k and p, is denoted as the boundary zone $\mathbf{BZ_{kp}}$.

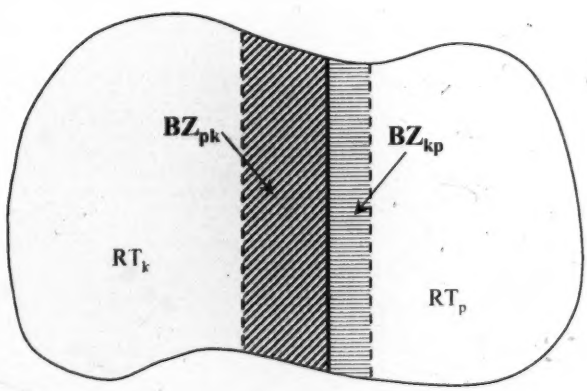

Figure 3.2: Boundary zone $\mathbf{BZ_{pk}}$ and $\mathbf{BZ_{kp}}$ corresponding to the area of influence of rock type p into rock type k and vice versa.

Precedence Rules

A set of precedence rules is needed to identify the predominant non-stationary factor at a boundary where more than two rock types converge. In Figure 3.3, a three rock type spatial arrangement is shown where precedence rules are used to establish the precedent rock types in the region A and B: in region A of $\mathbf{RT_k}$, $\mathbf{RT_p}$ is precedent over $\mathbf{RT_q}$, while in region B of $\mathbf{RT_p}$, $\mathbf{RT_q}$ is precedent over $\mathbf{RT_k}$.

Although the behavior of a property near such a boundary, could be explained by the overlapping of different geological controls, the task of identifying the individuals effects of each rock type and their interactions can be quite difficult. Geological properties are not usually additive and therefore the response of a combination of different rock types is complex.

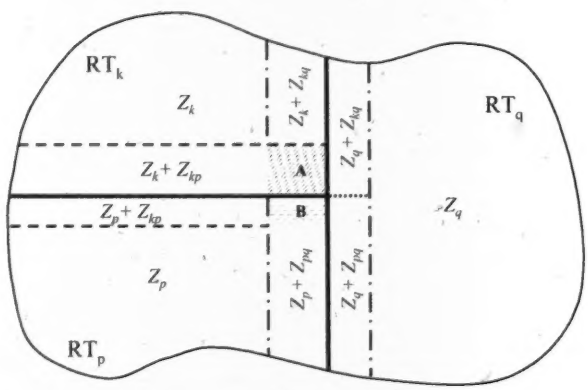

Figure 3.3: Three rock type example where a predefined precedence rule is used to determine the precedent rock type over two possibilities. In the region A of $\mathbf{RT_k}$, $\mathbf{RT_p}$ is precedent over $\mathbf{RT_q}$, while in region B of $\mathbf{RT_p}$, $\mathbf{RT_q}$ is precedent over $\mathbf{RT_k}$.

The precedence rules should be built by the modeler using all geological information available. The relative timing of intrusion, deposition or mineralisation events, geochemistry response of the protolith to an alteration or mineralisation process could be used to resolve timing and important variables. If the geological data do not provide sufficient information to establish a geological order of events, a neutral arrangement can be chosen. In this case, the precedent rock type p at a location $\mathbf{u}$ will be the one to which the distance to the boundary is the minimum over all surrounding rock types, that is, $\min\{d_{pk}(\mathbf{u})\}$ with $p=1,\ldots,K-1$ and $\mathbf{u}$ belongs to $\mathbf{BZ_{pk}}$.

For the same example illustrated in Figure 3.3, the result of a neutral arrangement is shown in Figure 3.4.

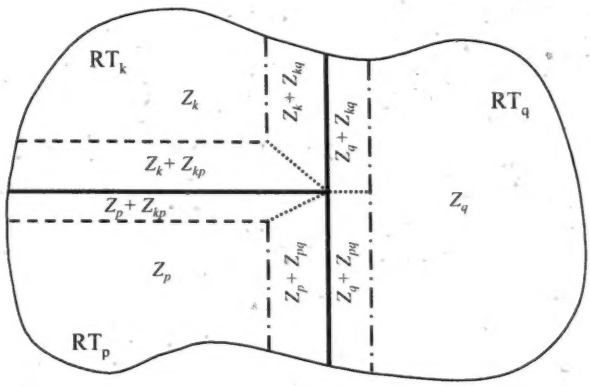

Figure 3.4: Three-rock type example where a neutral arrangement is chosen to define the shape of the zones of influence where two boundaries converge.

Statistical Parameters

The continuous random function, $Z(\mathbf{u})$ that represents the variable of interest over the study area, can be decomposed into stationary and non-stationary random variables by the rock type model, as shown in Equation 3.1. By definition, the non-stationary variable will take values only for locations within a distance to the boundary: $d_{pk}(\mathbf{u}) \leq dmax_{pk}$. A one dimensional example of how the random variable is decomposed into stationary and non-stationary components is presented in Figure 3.5.

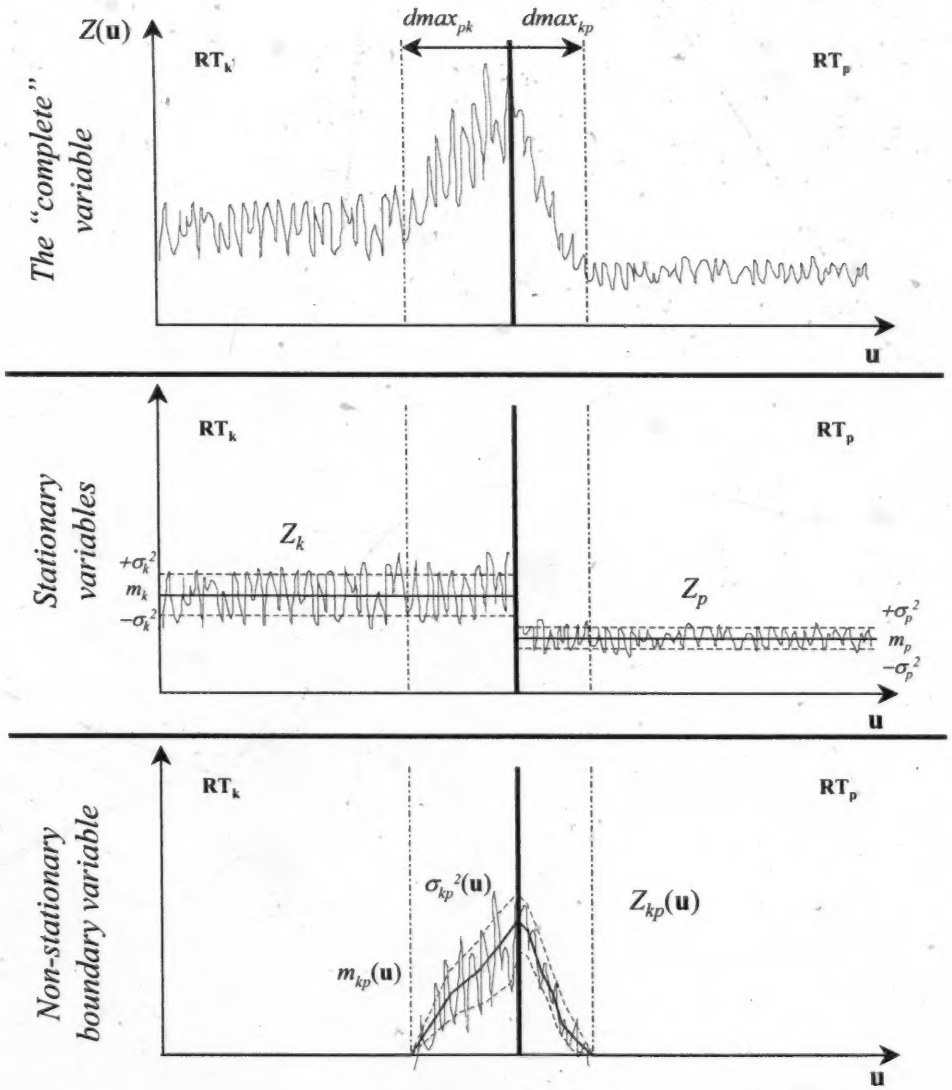

Figure 3.5: Decomposition of the random function $Z(\mathbf{u})$ in two stationary variables Z_k and Z_p, with constant mean and variance, and a non-stationary boundary variable $Z_{kp}(\mathbf{u})$, with a mean and variance that are functions of the distance to the boundary.

Mean

The mean function of the continuous random function $Z(\mathbf{u})$ for a specific rock type k will be the mean of the stationary variable Z_k plus the mean of any corresponding non-stationary variable $Z_{kp}(\mathbf{u})$:

$$E\{Z(\mathbf{u}_i)\} = E\{Z_k\} + E\{Z_{kp}(\mathbf{u}_i)\} = m_k + m_{kp}(\mathbf{u}_i) \quad \text{where } \mathbf{u}_i \in \mathbf{RT_k} \qquad (3.2)$$

where p is the adjacent rock type that shares a boundary with rock type k.

The stationary component of the mean (m_k) is independent of location and is a constant value. The non-stationary component of the mean (m_{kp}) is a function of the distance to the boundary, $d_{pk}(\mathbf{u})$ and takes values different than zero for locations within the boundary zone $\mathbf{BZ_{pk}}$:

$$m_{kp}(d_{pk}(\mathbf{u}_i)) = \begin{cases} 0 & , \text{if } d_{pk}(\mathbf{u}_i) \geq dmax_{pk} \\ f(d_{pk}(\mathbf{u}_i)) & , \text{otherwise} \end{cases} \quad \text{where } \mathbf{u}_i \in \mathbf{RT_k}$$

The non-stationary mean could be, for example, approximated by a linear function as we have chosen for the implementation presented later (see Chapter 4). There may be cases where a different approximation could be better. The methodology presented here could easily be adapted.

Rewriting Equation 3.2, the mean of rock type k in the presence of P non-stationary boundaries is:

$$E\{Z(\mathbf{u}_i)\} = \begin{cases} m_k & , \text{if } d_{pk}(\mathbf{u}_i) \geq dmax_{pk} \\ m_k + f(d_{pk}(\mathbf{u}_i)) & , \text{otherwise} \end{cases} \quad \text{where } \mathbf{u}_i \in \mathbf{RT_k} \qquad (3.3)$$

Variance

Similarly, the variance of $Z(\mathbf{u})$ for rock type k will be the sum of the stationary variance due to Z_k and the independent non-stationary variance due to $Z_{kp}(\mathbf{u})$ for the p adjacent rock type to rock type k, that is,

$$E\left\{\left(Z(\mathbf{u}_i) - E\{Z(\mathbf{u}_i)\}\right)^2\right\} = {\sigma_k}^2 + {\sigma_{kp}}^2(\mathbf{u}_i) \quad \text{where } \mathbf{u}_i \in \mathbf{RT_k} \qquad (3.4)$$

Since Z_k and $Z_{kp}(\mathbf{u})$ are independent random variables, the cross terms of the squared product are zero;

$$E\left\{\left(Z(\mathbf{u}_i)-E\{Z(\mathbf{u}_i)\}\right)^2\right\}=E\left\{\left[\left(Z_k(\mathbf{u}_i)+Z_{kp}(\mathbf{u}_i)\right)-\left(m_k+m_{kp}(\mathbf{u}_i)\right)\right]^2\right\}$$

$$=E\left\{\left(Z_k(\mathbf{u}_i)+Z_{kp}(\mathbf{u}_i)\right)^2-2\left(Z_k(\mathbf{u}_i)+Z_{kp}(\mathbf{u}_i)\right)\cdot\left(m_k+m_{kp}(\mathbf{u}_i)\right)+\left(m_k+m_{kp}(\mathbf{u}_i)\right)^2\right\}$$

$$=E\left\{Z_k(\mathbf{u}_i)^2\right\}+2E\left\{Z_k(\mathbf{u}_i)\cdot Z_{kp}(\mathbf{u}_i)\right\}+E\left\{Z_{kp}(\mathbf{u}_i)^2\right\}-m_k{}^2-2m_k\cdot m_{kp}(\mathbf{u}_i)-m_{kp}(\mathbf{u}_i)^2$$

$$=E\left\{Z_k(\mathbf{u}_i)^2\right\}+2E\left\{Z_k(\mathbf{u}_i)\right\}\cdot E\left\{Z_{kp}(\mathbf{u}_i)\right\}+E\left\{Z_{kp}(\mathbf{u}_i)^2\right\}-m_k{}^2-2m_k\cdot m_{kp}(\mathbf{u}_i)-m_{kp}(\mathbf{u}_i)^2$$

$$=E\left\{Z_k(\mathbf{u}_i)^2\right\}-m_k{}^2+E\left\{Z_{kp}(\mathbf{u}_i)^2\right\}-m_{kp}(\mathbf{u}_i)^2$$

$$=\sigma_k{}^2+\sigma_{kp}{}^2(\mathbf{u}_i)$$

As with the mean, the stationary component of the variance ($\sigma_k{}^2$) is independent of location and is a constant value. The non-stationary component of the variance ($\sigma_{kp}{}^2$) is a function of the distance to the boundary. Beyond the maximum distance of influence it is equal to zero. Within the boundary zone $\mathbf{BZ_{pk}}$ it is defined as:

$$\sigma_{kp}{}^2(d_{pk}(\mathbf{u}_i))=\begin{cases}0 & \text{, if } d_{pk}(\mathbf{u}_i)\geq dmax_{pk}\\ g(d_{pk}(\mathbf{u}_i)) & \text{, otherwise}\end{cases} \quad \text{where } \mathbf{u}_i\in\mathbf{RT_k}$$

$g(d_{pk}(\mathbf{u}_i))$ was chosen as a linear function in the following examples and implementation.

A summarized expression for the variance of a random function $Z(\mathbf{u})$ in a rock type k with P non-stationary boundaries is:

$$E\left\{\left(Z(\mathbf{u}_i)-E\{Z(\mathbf{u}_i)\}\right)^2\right\}=\begin{cases}\sigma_k{}^2 & \text{, if } d_{pk}(\mathbf{u}_i)\geq dmax_{pk}\\ \sigma_k{}^2+g(d_{pk}(\mathbf{u}_i)) & \text{, otherwise}\end{cases} \quad \text{where } \mathbf{u}_i\in\mathbf{RT_k}$$

Covariance

As with the mean and variance, the covariance structure between two rock types that share a local non-stationary boundary consists of: a stationary and a non-stationary component.

$$Cov_Z(\mathbf{u}_i,\mathbf{v}_i) = E\left\{\left(Z(\mathbf{u}_i)-m(\mathbf{u}_i)\right)\cdot\left(Z(\mathbf{v}_i)-m(\mathbf{v}_i)\right)\right\} = Cov_Z^{S}(\mathbf{h}) + Cov_Z^{NS}(\mathbf{u}_i,\mathbf{v}_i)$$

where $\mathbf{h}=\mathbf{u}_i - \mathbf{v}_i$.

Since Z_k and $Z_{kp}(\mathbf{u})$ are independent random variables, the cross terms are zero, therefore the covariance of $Z(\mathbf{u})$ is the sum of the stationary and non-stationary components. The combination of these components corresponds to a local linear model of coregionalization. In some respects this model is one of regionalization and not coregionalization; however it does involve different variables (grades in different rock types), therefore, the terms coregionalization is used.

The stationary component of the covariance can be calculated and modeled from data pairs within the same internal stationary portion of a rock type, that is $\mathbf{u}_i$ and $\mathbf{v}_i$ belong to rock type k, and do not belong to any $\mathbf{BZ_{pk}}$.

The inference of the covariance structure will be possible provided a reasonable number of pairs exist within the stationary and non-stationary regions.

To obtain the non-stationary component of the covariance model we will assume that the shape, anisotropies and relative nugget effect of the correlation for the non-stationary variable $Z_{kp}(\mathbf{u})$ $k,p=1,\ldots,K$ are stationary and that they can be specified by the modeler. But due to the non-stationary nature of variable $Z(\mathbf{u})$ at the boundary zone, this stationary spatial model shape has to by scaled at each point by a non-stationary mean and variance. The relative standardized variogram model for the boundary zone, $\mathbf{BZ_{pk}}$, corresponding to the stationary shape is:

$$\hat{\gamma}_{kp}(\mathbf{u}_i,\mathbf{v}_i) = \frac{1}{2}\cdot E\left\{\left[\frac{Z(\mathbf{u}_i)-(m_{kp}(\mathbf{u}_i)+m_k)}{(\sigma_{kp}(\mathbf{u}_i)+\sigma_k)} - \frac{Z(\mathbf{v}_i)-(m_{kp}(\mathbf{v}_i)+m_k)}{(\sigma_{kp}(\mathbf{v}_i)+\sigma_k)}\right]^2\right\}$$

The stationary component of the mean, in the previous expression can be either m_k or m_p depending whether $\mathbf{u}_i$ or $\mathbf{v}_i$ belongs to rock type k or rock type p. The same occurs for the stationary component of the variance.

Expanding the squared difference we obtained that:

$$2\hat{\gamma}_{kp}(\mathbf{u}_i,\mathbf{v}_i) = E\left\{\begin{array}{l} \dfrac{Z(\mathbf{u}_i)^2 - 2Z(\mathbf{u}_i)\cdot m(\mathbf{u}_i) + m(\mathbf{u}_i)^2}{\sigma(\mathbf{u}_i)^2} \\ -2\left(\dfrac{Z(\mathbf{u}_i)\cdot Z(\mathbf{v}_i) - Z(\mathbf{v}_i)\cdot m(\mathbf{u}_i) - Z(\mathbf{u}_i)\cdot m(\mathbf{v}_i) + m(\mathbf{u}_i)\cdot m(\mathbf{v}_i)}{\sigma(\mathbf{u}_i)\cdot\sigma(\mathbf{v}_i)}\right) \\ +\dfrac{Z(\mathbf{v}_i)^2 - 2Z(\mathbf{v}_i)\cdot m(\mathbf{v}_i) + m(\mathbf{v}_i)^2}{\sigma(\mathbf{v}_i)^2} \end{array}\right\}$$

To simplify the notation the mean and variance of $\mathbf{u}_i$ and $\mathbf{v}_i$ has been replaced by:

$$m(\mathbf{u}_i) = m_{kp}(\mathbf{u}_i) + m_k \quad \textit{if } \mathbf{u}_i \in \mathbf{RT}_k \quad \textit{or} \quad m(\mathbf{u}_i) = m_{kp}(\mathbf{u}_i) + m_p \quad \textit{if } \mathbf{u}_i \in \mathbf{RT}_p$$
$$\sigma(\mathbf{u}_i) = \sigma_{kp}(\mathbf{u}_i) + \sigma_k \quad \textit{if } \mathbf{u}_i \in \mathbf{RT}_k \quad \textit{or} \quad \sigma(\mathbf{u}_i) = \sigma_{kp}(\mathbf{u}_i) + \sigma_p \quad \textit{if } \mathbf{u}_i \in \mathbf{RT}_p$$

and

$$m(\mathbf{v}_i) = m_{kp}(\mathbf{v}_i) + m_k \quad \textit{if } \mathbf{v}_i \in \mathbf{RT}_k \quad \textit{or} \quad m(\mathbf{v}_i) = m_{kp}(\mathbf{v}_i) + m_p \quad \textit{if } \mathbf{v}_i \in \mathbf{RT}_p$$
$$\sigma(\mathbf{v}_i) = \sigma_{kp}(\mathbf{v}_i) + \sigma_k \quad \textit{if } \mathbf{v}_i \in \mathbf{RT}_k \quad \textit{or} \quad \sigma(\mathbf{v}_i) = \sigma_{kp}(\mathbf{v}_i) + \sigma_p \quad \textit{if } \mathbf{v}_i \in \mathbf{RT}_p$$

Since $E\{Z(\mathbf{u}_i)^2\} = \sigma(\mathbf{u}_i)^2 + m(\mathbf{u}_i)$ and $E\{Z(\mathbf{u}_i)\} = m(\mathbf{u}_i)$, we can simplify the previous expression as follows,

$$\begin{aligned} 2\hat{\gamma}_{kp}(\mathbf{u}_i,\mathbf{v}_i) &= \frac{\sigma(\mathbf{u}_i)^2 + m(\mathbf{u}_i)^2 - 2\cdot m(\mathbf{u}_i)^2 + m(\mathbf{u}_i)^2}{\sigma(\mathbf{u}_i)^2} \\ &\quad -2\left(\frac{E\{Z(\mathbf{u}_i)\cdot Z(\mathbf{v}_i)\} - m(\mathbf{v}_i)\cdot m(\mathbf{u}_i) - m(\mathbf{u}_i)\cdot m(\mathbf{v}_i) + m(\mathbf{u}_i)\cdot m(\mathbf{v}_i)}{\sigma(\mathbf{u}_i)\cdot\sigma(\mathbf{v}_i)}\right) \\ &\quad +\frac{\sigma(\mathbf{v}_i)^2 + m(\mathbf{v}_i)^2 - 2 + m(\mathbf{v}_i)^2 + m(\mathbf{v}_i)^2}{\sigma(\mathbf{v}_i)^2} \\ &= 2 - 2\left(\frac{E\{Z(\mathbf{u}_i)\cdot Z(\mathbf{v}_i)\} - m(\mathbf{v}_i)\cdot m(\mathbf{u}_i)}{\sigma(\mathbf{u}_i)\cdot\sigma(\mathbf{v}_i)}\right) \end{aligned}$$

$$\hat{\gamma}_{kp}(\mathbf{u}_i,\mathbf{v}_i) = 1 - \frac{Cov_c^{NS}(\mathbf{u}_i,\mathbf{v}_i)}{\sigma(\mathbf{u}_i)\cdot\sigma(\mathbf{v}_i)}$$

From this it is clear that the covariance must be scaled by the non-stationary standard deviations of the $\mathbf{u}_i$ and $\mathbf{v}_i$ locations. These standard deviations are a function of the distance to the boundary. Reordering the terms, we can obtain an expression for the non-stationary covariance model component:

$$Cov_Z^{NS}(\mathbf{u}_i,\mathbf{v}_i) = E\{Z(\mathbf{u}_i)\cdot Z(\mathbf{v}_i)\} - (m_{kp}(\mathbf{v}_i)+m_k)\cdot(m_{kp}(\mathbf{u}_i)+m_k)$$
$$= (1-\hat{\gamma}_{kp}(\mathbf{u}_i,\mathbf{v}_i))\cdot(\sigma_{kp}(\mathbf{u}_i)+\sigma_k)\cdot(\sigma_{kp}(\mathbf{v}_i)+\sigma_k) \quad (3.6)$$

Since the shape, anisotropies and nugget effect of the relative standardize variogram are inputs from the modeler, the only parameter that must be established for the non-stationary covariance model is the range. A more general formalism could be considered in the future.

Depending whether both $\mathbf{u}_i$ and $\mathbf{v}_i$, only one of them or neither of them, fall at the boundary zone $\mathbf{BZ_{pk}}$, the covariance structure of $Z(\mathbf{u})$ will be sum of the stationary and non-stationary components or only the stationary component (Figure 3.6)

$$Cov_Z(\mathbf{u}_i,\mathbf{v}_i) = \begin{cases} Cov_Z^S(\mathbf{h}) & ,if\ \mathbf{u}_i \in \mathbf{RT_k}, \notin \mathbf{BZ_{pk}} and\ \mathbf{v}_i \in \mathbf{RT_k}, \notin \mathbf{BZ_{pk}} \\ Cov_Z^S(\mathbf{h}) & ,if\ \mathbf{u}_i \in \mathbf{RT_k}, \notin \mathbf{BZ_{pk}} and\ \mathbf{v}_i \in \mathbf{RT_k}, \in \mathbf{BZ_{pk}} \\ Cov_Z^S(\mathbf{h}) + Cov_Z^{NS}(\mathbf{u}_i,\mathbf{v}_i) & ,if\ \mathbf{u}_i \in \mathbf{RT_k}, \in \mathbf{BZ_{pk}} and\ \mathbf{v}_i \in \mathbf{RT_k}, \in \mathbf{BZ_{pk}} \\ Cov_Z^{NS}(\mathbf{u}_i,\mathbf{v}_i) & ,if\ \mathbf{u}_i \in \mathbf{RT_k}, \in \mathbf{BZ_{pk}} and\ \mathbf{v}_i \in \mathbf{RT_p}, \in \mathbf{BZ_{kp}} \\ 0 & ,if\ \mathbf{u}_i \in \mathbf{RT_k}, \notin \mathbf{BZ_{pk}} and\ \mathbf{v}_i \in \mathbf{RT_p}, \in \mathbf{BZ_{kp}} \\ 0 & ,if\ \mathbf{u}_i \in \mathbf{RT_k}, \notin \mathbf{BZ_{pk}} and\ \mathbf{v}_i \in \mathbf{RT_p}, \notin \mathbf{BZ_{kp}} \end{cases}$$

(3.5)

If the head and tail of the pair are in different rock types, the stationary component of the covariance structure is zero, since there is no correlation between the stationary variables Z_k and Z_p across the boundary.

Optimization of the Statistical Parameters

The mean, variance and covariance are crucial for estimation and simulation. We must fit the distribution of the random variable $Z(\mathbf{u})$ at the boundary zone knowing the stationary rock type, precedence rules and maximum distances of influence.

Although we know that the non-stationary behavior is a function of the distance of the sample to the boundary, there are several possible analytical expressions that fit the distribution of the non-stationary random variable. For the purpose of showing the proposed methodology we will consider that the non-stationary

component of mean and variance follow a linear function (Figure 3.7) of the distance to the boundary (d_{pk}). In this scenario, the optimization of the parameter m_{kp} and σ_{kp}^2 will be equivalent to optimizing estimates of the intercepts at distance zero from the boundary: a_{kp} and b_{kp}.

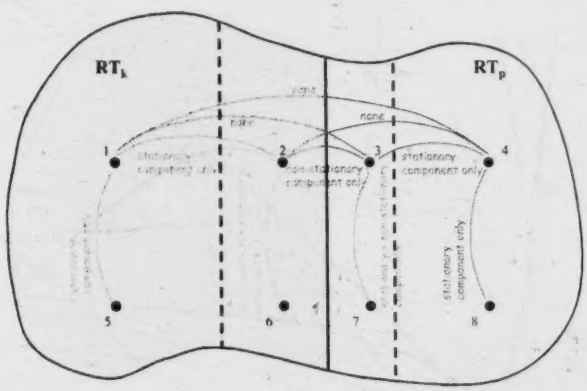

Figure 3.6: Two-rock type example showing how the stationary and non-stationary components of the covariance model are add together for different configurations.

Mean Optimization

The mean m_{kp} is optimized given that m_k is known from the experimental average of data within rock type k, outside any boundary zone. Using Equation 3.2, the objective function to be minimized is written:

$$O_m = \sum_{k=1}^{K} \sum_{p=1}^{P_k} \sum_{i=1}^{N_{kp}} \left[z(\mathbf{u}_i) - (\hat{m}_k + m_{kp}(\mathbf{u}_i)) \right]^2$$

where $z(\mathbf{u}_i)$ is the outcome value at every location $\mathbf{u}_i \in \mathbf{BZ_{pk}}$, $\hat{m}_k$ is the experimental average (Equation 3.7) and $m_{kp}(\mathbf{u}_i)$ is the non-stationary mean at location $\mathbf{u}_i$ derived from Equation 3.8. N_{kp} is the total number of data in $\mathbf{BZ_{pk}}$. The experimental mean of rock type k is calculated as:

$$\hat{m}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} w(\mathbf{u}_i) \cdot z_k(\mathbf{u}_i) \quad \forall \mathbf{u}_i \in \mathbf{RT_k}, \notin \mathbf{BZ_{kp}}; \forall p \qquad (3.7)$$

where N_k is all data within the internal stationary portion and $w(\mathbf{u}_i)$ are the declustering weights associated to the location $\mathbf{u}_i$.

Although the mean of rock type k should include all samples in this geological unit, the samples in any boundary zone are excluded since their non-stationary component is yet to be determined.

In Equation 3.7, N_k denotes the number of data within the internal stationary portion of rock type k, that is, outside any boundary zone with any adjacent rock type p.

$$m_{kp}(\mathbf{u}_i) = \begin{cases} \dfrac{\left(dmax_{kp} - d_{kp}(\mathbf{u}_i)\right)}{dmax_{km}} \cdot a_{kp} & \text{for } 0 \le d_{kp}(\mathbf{u}_i) \le dmax_{kp} \\ \dfrac{\left(dmax_{pk} - d_{pk}(\mathbf{u}_i)\right)}{dmax_{pk}} \cdot a_{kp} & \text{for } 0 \le d_{pk}(\mathbf{u}_i) \le dmax_{pk} \\ 0 & \text{for } d_{kp}(\mathbf{u}_i) \ge dmax_{kp} \text{ and } d_{pk}(\mathbf{u}_i) \ge dmax_{pk} \end{cases} \tag{3.8}$$

From Equation 3.8 it is clear that the optimization of m_{kp}, amounts to the optimization of a_{kp}, given $dmax_{kp}$ and $dmax_{pk}$, then the optimization of the mean can be achieved by iteratively modifying a_{kp} $\forall k,p$, in a random fashion while accepting all changes in a_{kp} that reduce objective function. This is a simplified version of the simulated annealing formalism, where only favorable perturbations are accepted. For our implementation, we will assume that $a_{kp}= a_{pk}$, but this hypothesis can easily be changed if a discontinuity at the boundary is clear.

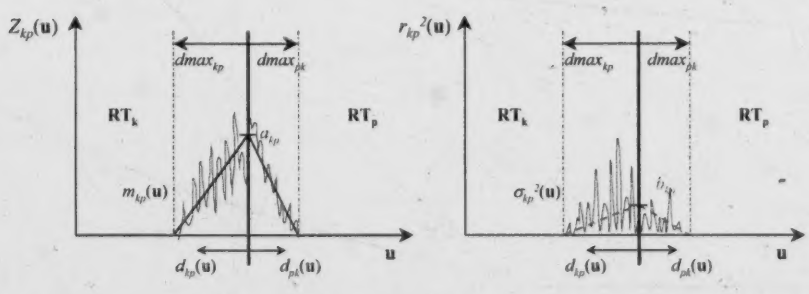

Figure 3.7: Mean and variance of the random variable Z_{kp} $(\mathbf{u})$, modeled by a linear function of the distance to the boundary.

Variance Optimization

Since the variance is a statistic of second order that depends on the mean, once the optimum of m_{kp} is found, we can proceed to optimize ${\sigma_{kp}}^2$ assuming the ${\sigma_k}^2$, $k=1,...,K$ values are known from the experimental variance of data within the internal stationary portion of rock type k.

From Equation 3.4 the optimum ${\sigma_{kp}}^2$, will be the one that minimizes the following objective function:

$$O_{\sigma^2} = \sum_{k=1}^{K}\sum_{p=1}^{P}\sum_{i=1}^{N_{kp}}\left[\mathrm{r}(\mathbf{u}_i)^2 - (\hat{\sigma}_k^{\,2} + {\sigma_{kp}}^2(\mathbf{u}_i))\right]^2 \qquad (3.9)$$

where $\mathrm{r}(\mathbf{u}_i)$ is the residual value at every location $\mathbf{u}_i \in \mathbf{BZ_{pk}}$, that is, $r(\mathbf{u}_i) = z(\mathbf{u}_i) - (m_k + m_{kp}(\mathbf{u}_i))$. ${\hat{\sigma}_k}^2$ is the experimental variance (Equation 3.10), and ${\sigma_{kp}}^2(\mathbf{u}_i)$ is the non-stationary variance at location $\mathbf{u}_i$ derived from Equation 3.11. N_{kp} is total number of data in $\mathbf{BZ_{pk}}$. The experimental variance of rock type k is calculated as:

$$\hat{\sigma}_k^{\,2} = \frac{1}{N_k}\sum_{i=1}^{N_k} w(\mathbf{u}_i)\cdot r_k(\mathbf{u}_i)^2 \quad \forall \mathbf{u}_i \in \mathbf{RT_k}, \notin \mathbf{BZ_{kp}}; \forall p \qquad (3.10)$$

where N_k is all data within the internal stationary portion and $w(\mathbf{u}_i)$ are the declustering weights associated to the location $\mathbf{u}_i$.

As with the mean, the optimization of ${\sigma_{kp}}^2$ is the same as the optimization of b_{kp}, given $dmax_{kp}$ and $dmax_{pk}$. We assume that $b_{kp} = b_{pk}$.

$${\sigma_{kp}}^2(\mathbf{u}_i) = \begin{cases} \dfrac{\left(dmax_{kp} - d_{kp}(\mathbf{u}_i)\right)}{dmax_{kp}}\cdot b_{kp} & \text{for } 0 \le d_{kp}(\mathbf{u}_i) \le dmax_{kp} \\ \dfrac{\left(dmax_{pk} - d_{pk}(\mathbf{u}_i)\right)}{dmax_{pk}}\cdot b_{kp} & \text{for } 0 \le d_{pk}(\mathbf{u}_i) \le dmax_{pk} \\ 0 & \text{for } d_{kp}(\mathbf{u}_i) \ge dmax_{kp} \text{ and } d_{pk}(\mathbf{u}_i) \ge dmax_{pk} \end{cases} \qquad (3.11)$$

Covariance Optimization

Establishing the covariance model of a boundary zone is equivalent to optimizing the range of the relative standardized variogram that represents the stationary shape that is scaled by the standard deviations of the paired samples (Figure 3.8). This assumption provides some advantages over a full optimization algorithm to find the non-stationary covariance structure. Simplicity and fewer artifacts are the main advantages.

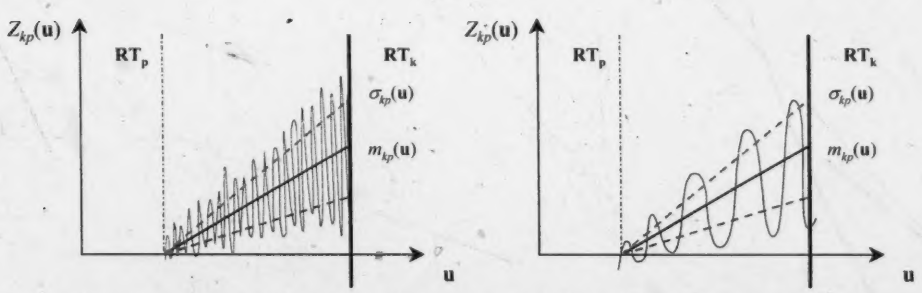

Figure 3.8: The non-stationary covariance of $Z_{kp}(\mathbf{u})$ is defined by its non-stationary mean and variance and the shape of the correlation, which for the same mean and variance can be different as represented by the outcomes profile. The range of correlation is higher in the right.

To find the optimum range, we need to minimize the following objective function:

$$O_{Cov} = \sum_{i=1}^{N} \left[\hat{C}\left(z(\mathbf{u}_i), z(\mathbf{v}_i)\right) - C_{MOD}\left(z(\mathbf{u}_i), z(\mathbf{v}_i)\right) \right]^2$$

where $\hat{C}$ denotes the experimental covariance of the pair located at $\mathbf{u}_i$ and $\mathbf{v}_i$, which is just the multiplication of the two residual values:

$$\hat{C}\left(z(\mathbf{u}_i), z(\mathbf{v}_i)\right) = r(\mathbf{u}_i) \cdot r(\mathbf{v}_i)$$

and C_{MOD} the modeled boundary covariance, which is the sum of the stationary and non-stationary component.

The squared difference between the experimental and modeled covariance is evaluated for randomly changed range that is iteratively modified until this difference is minimized. Since the anisotropies are fixed during the optimization

process, the ratios between the major and minor horizontal ranges and the major and vertical ranges are known. Consequently, we only need to find the total range.

More parameters could be optimized in the same fashion, although we do not recommend a full optimization without many data, since this could lead to artifacts.

Estimation in presence of local non-stationary boundaries

Once the mean, variance and covariance model have been determined both for the stationary and non-stationary regions of a geological model, estimation can proceed. It is propose a non-stationary form of simple cokriging. The basic linear regression estimator of kriging is:

$$z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_\alpha(\mathbf{u}) \cdot \left[z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha) \right] \qquad (3.12)$$

where $z^*(\mathbf{u})$ is the estimate at unsampled location $\mathbf{u}$, $m(\mathbf{u})$ is the mean value at location $\mathbf{u}$, $\lambda_\alpha(\mathbf{u})$ is the weight assigned to datum $z(\mathbf{u}_\alpha)$, n are the closest data to the location $\mathbf{u}$ being estimated, and $m(\mathbf{u}_\alpha)$ are the n mean values at the data locations. In this case, the expected values for the mean at the location being estimated and at each data location, will be calculated from Equation 3.3, that is, if any of the closest data or the location being estimated is within the zone of influence of a boundary, its mean will include a non-stationary component that will be calculated from a function of the distance to the boundary.

To find the optimal weights $\lambda_\alpha(\mathbf{u})$, $\alpha=1,\ldots,n$ the following matrix system (the kriging equations) must be solved:

$$\sum_{\beta=1}^{n} \lambda_\beta(\mathbf{u}) \cdot Cov(\mathbf{u}_\alpha, \mathbf{u}_\beta) = Cov(\mathbf{u}, \mathbf{u}_\alpha) \quad \text{with } \alpha, \beta = 1,\ldots,n$$

where $\lambda_\alpha(\mathbf{u})$, $\alpha=1,..,n$ are the simple kriging weights, $Cov(\mathbf{u}_\alpha, \mathbf{u}_\beta)$, $\alpha,\beta=1,..,n$ correspond to the data-to-data covariances, and $Cov(\mathbf{u}, \mathbf{u}_\alpha)$, $\alpha=1,..,n$ are the data-to-estimate covariances, which in matrix notation is equivalent to:

$$\begin{bmatrix} Cov(\mathbf{u}_1,\mathbf{u}_1) & \cdots & Cov(\mathbf{u}_1,\mathbf{u}_n) \\ \vdots & \ddots & \vdots \\ Cov(\mathbf{u}_n,\mathbf{u}_1) & \cdots & Cov(\mathbf{u}_n,\mathbf{u}_n) \end{bmatrix} \cdot \begin{bmatrix} \lambda_1(\mathbf{u}) \\ \vdots \\ \lambda_n(\mathbf{u}) \end{bmatrix} = \begin{bmatrix} Cov(\mathbf{u},\mathbf{u}_1) \\ \vdots \\ Cov(\mathbf{u},\mathbf{u}_n) \end{bmatrix} \tag{3.13}$$

The kriging weights are obtained by solving this system of linear equations. This system has a solution and is unique if the covariance model $Cov_Z(\mathbf{h})$ is a legitimate model, that is, was fit with a positive definite function, and data are not collocated.

For a geological model with local non-stationary boundaries, the data covariance matrix and the vector of data-to-estimate covariances are obtained by combining the stationary and non-stationary covariance model components as in Equation 3.5. The stationary covariance model is obtained from fitting the experimental variogram calculated with the data within the stationary regions within each rock type, while the non-stationary covariance component is calculated from Equation 3.6, using the correlation spatial model obtained from the optimization process describe in the previous section. The methodology and calculations are shown in more detail in an example in the next section.

Simple kriging estimator is unbiased and provides the minimum estimation variance estimate,

$$\sigma_E^2(\mathbf{u}) = \sigma^2(\mathbf{u}) - \sum_{\alpha=1}^{n} \lambda_\alpha(\mathbf{u}) \cdot Cov(\mathbf{u},\mathbf{u}_\alpha) \tag{3.14}$$

where $\sigma^2(\mathbf{u})$ is the variance, which in our case has a stationary and a non-stationary component. Kriging also accounts for redundancy between the data and closeness of the data to where we are estimating. Spatial continuity is provided by the variogram model. The greatest disadvantage of kriging is that the estimates are smooth and the joint variability of the kriged estimates is incorrect.

The amount of missing variability is the kriging variance (Equation 3.14). Sequential Gaussian simulation overcomes this problem by adding back this missing variability, as it adds a random residual to the estimate, drawn from a normal distribution with zero mean and variance equal to the kriging variance. In the case described in this thesis, the estimation variance has also a non-stationary component that makes the implementation of sequential Gaussian simulation in the presence of local non-stationary boundaries delicate. This implementation is part of the proposed future work.

Chapter 4

Implementation Details

RT model and boundaries

The first step is the correct labeling of every location with the corresponding rock type and relevant closest precedent boundary zone and its distance to the boundary. To help with this task a small program coded in FORTRAN called boundmod was created.

The input data is a grid-type file of the geological model with the same format as the one used in GSLIB (Deutsch and Journel, 1998), the number of rock types available in the model, the matrix with all maximum distances of influence between rock types, and the set of precedence rules. The output file is a grid file with 3 variables: the rock type (from input), the boundary zone identified by the rock type code of the precedent influencing rock type at that location or a default value if is beyond any maximum distance of influence (internal stationary regions), and the distance to the boundary (measured from the center of the node) with the precedent rock type.

In summary, the program loops over all nodes of the geological model grid, and checks within a search window if different rock types exist within the surrounding nodes. The cell may be assigned with a neighbor rock type if a boundary zone is defined for the two codes and the precedence rules are met. A more detailed structure of the program is shown in the flow chart of Figure 4.1.

The parameter file to run this program (Figure 4.2) has 4 major parts. The first one consists of four lines: the name of the file of the geological model, the variable number that contain the rock type codes at each node, the grid definition and the number of different rock types. In order for this and the following programs to work properly, the rock types codes must be successive integers, starting at 1. The grid definition is given by the number of nodes, coordinate of

the origin (located at the center of the cell) and size of the cell for each axis: X, Y and Z.

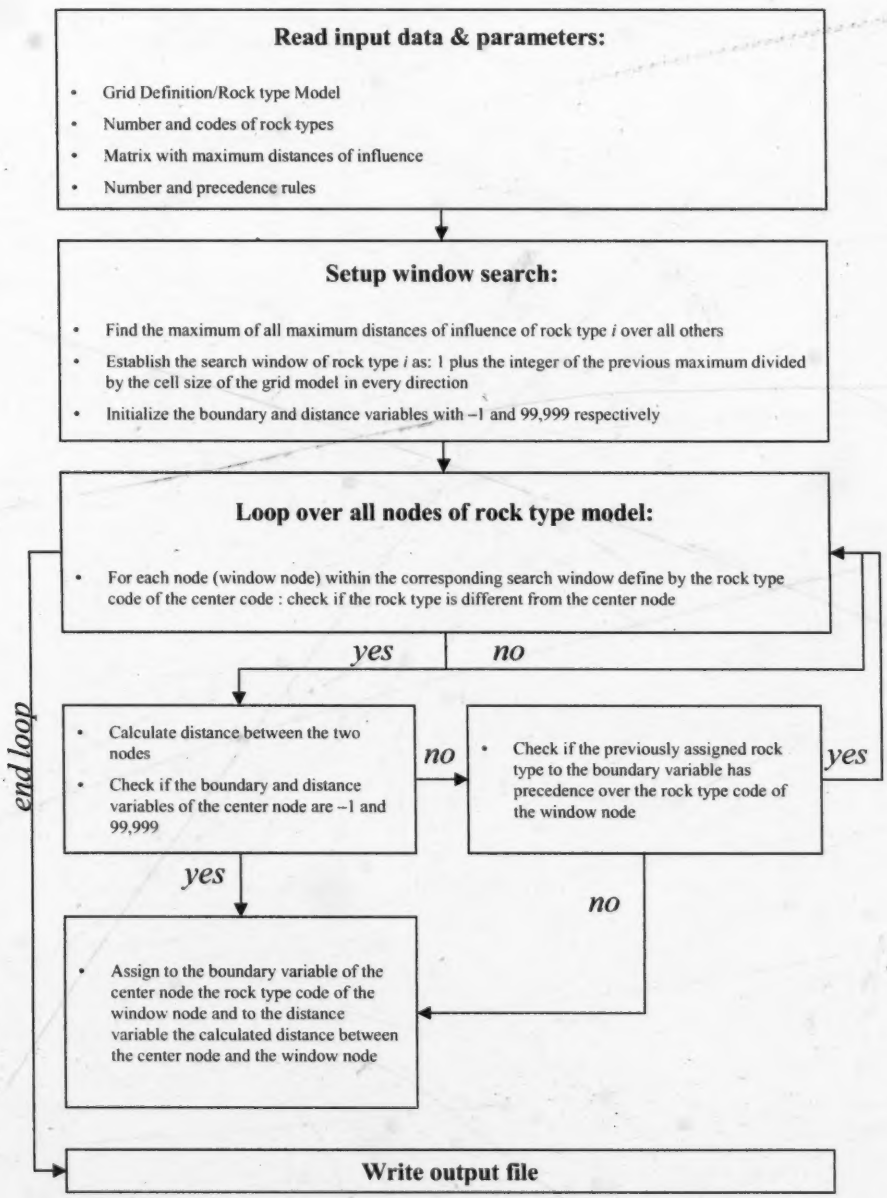
Read input data & parameters:
Grid Definition/Rock type Model
Number and codes of rock types
Matrix with maximum distances of influence
Number and precedence rules
Setup window search:
Find the maximum of all maximum distances of influence of rock type i over all others
Establish the search window of rock type i as: 1 plus the integer of the previous maximum divided by the cell size of the grid model in every direction
Initialize the boundary and distance variables with –1 and 99,999 respectively
Loop over all nodes of rock type model:
For each node (window node) within the corresponding search window define by the rock type code of the center code : check if the rock type is different from the center node
yes
no
end loop
Calculate distance between the two nodes
Check if the boundary and distance variables of the center node are –1 and 99,999
no
Check if the previously assigned rock type to the boundary variable has precedence over the rock type code of the window node
yes
yes
no
Assign to the boundary variable of the center node the rock type code of the window node and to the distance variable the calculated distance between the center node and the window node
Write output file

Figure 4.1: Flow chart of the program boundmod.

The second part corresponds to the matrix of maximum distances of influence and it will have as many lines as different rock types in the model. Each line corresponds to the maximum distance of influence of that rock type over all others. For example in, the parameter file shown in Figure 4.2, the maximum distance of influence of rock type *2* into rock type *1*, is 10 meters, i.e., $dmax_{21}$=10. The maximum distance of influence of $\mathbf{RT_2}$ with itself is denoted by, -1.0, a default number. If two rock types are not in contact with each other or a 'hard' boundary between them is assumed, the maximum distance of influence is set to zero, as in $dmax_{23}$=0.0.

The third part of the parameter file corresponds to specification of the precedence rules: the number of them and the rules expressed as pairs of rock type codes. The order of the pairs is from oldest to youngest, or from less to more precedent, and is given by: "rock type code – boundary code". In the program, at each grid node, the array with all precedence rules is checked in order, for both pairs: the assigned boundary code in a previous iteration and the rock type code plus the possible boundary value obtained from the window node in the current iteration. If the pair is found within the list of pairs, the correlative number of the rule is assigned to the pair. At the beginning of this check, both pairs are assigned with a correlative number of zero. The boundary code obtained from the window node overwrites a previous value of the boundary variable only if the rule number assigned to this pair is greater than the rule number of the previously assigned pair. In the parameter file shown in the example, the pair 2-4, rule number 4, is precedent over all previous pairs in the list: 3-1, 4-1 and 1-4. Leaving the number of precedence rules to zero is equivalent to choosing a neutral arrangement where the precedent rock type is chosen as the one with the minimum distance to a boundary.

Finally the last line and part of the parameter file corresponds to the name that will be given to the output file of the program.

```
          Parameters for BOUNDMOD
          ***********************

START OF PARAMETERS:
../data/rtmodel.out                  -input file with rock type model
1                                    -column with RT value
100    0.5      1.0                  -nx,xmn,xsiz
100    0.5      1.0                  -ny,ymn,ysiz
1      0.0      1.0                  -nz,zmn,zsiz
3                                    -number of rock types (correlative starting at 1)
-1.0   50.0     20.0                 -maximum distance of influence of RT 1 into RT j
10.0   -1.0     50.0                 -maximum distance of influence of RT 2 into RT j
20.0   30.0     -1.0                 -maximum distance of influence of RT 3 into RT j
2                                    -number of precendence rules
1 3                                  -pair RTi,RTj, with the oldest precedence
3 1                                  -pair RTi,RTj, with the youngest precedence
output.out                           -output file
```

Figure 4.2: Parameter file for boundmod program.

The following example illustrates the results of this program for a synthetic rock type model. The input rock type model has 4 rock types with a spatial configuration as shown in Figure 4.3, which reflects two sedimentary or volcanic strata cut by a reverse fault. The maximum distances of influence, are given in the parameter file of Figure 4.2, and reflect a broader influence across the fault than across the horizontal contact between the two strata, with small differences between the upper and lower strata and the footwall and hanging walls of the fault.

For an exercise with 4 precedence rules (Figure 4.4), where the boundary zones generated by fault are precedent over the ones generate by the horizontal contact between layers, the program outputs for boundary zones and distance to boundary are shown in Figure 4.5. If a neutral arrangement is chosen, which is equivalent to set the number of precedence rules to zero, the boundary zone and distance to the boundary are assigned based on the closest boundary code an its distance to the node (Figure 4.6).

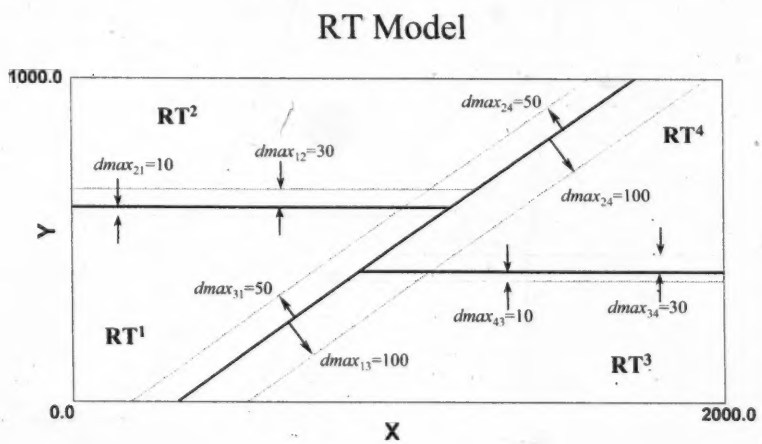

Figure 4.3: Example of a geological model with 4 rock types and maximum distances of influence define for each boundary. The influence of the fault has a larger extent compared to the influence of the stratigraphic contact.

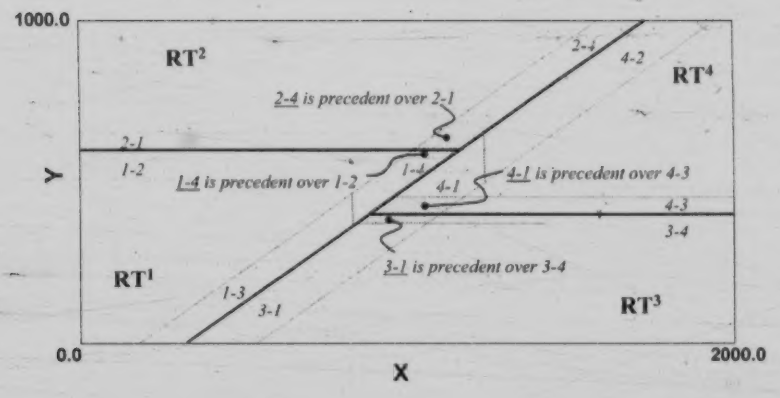

Figure 4.4: Set of four precedence rules: *3-1*, *4-1*, *1-4* and *2-4*, for the example of Figure 4.3. With these precedence rules in the hatched regions the boundary zones defined by the fault are predominant over the boundary zones defined by the horizontal contact between stratigraphic layers.

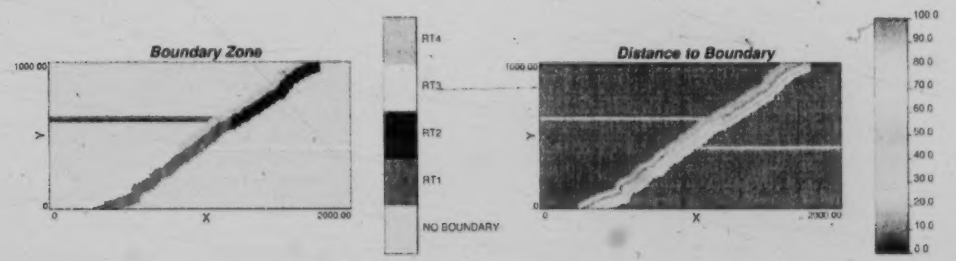

Figure 4.5: Output of the boundmod program for the example of Figure 4.3, considering a set of 4 precedence rules. The left map corresponds to the categorical variable boundary zone, a default code (NO BOUNDARY) is given to locations beyond the corresponding maximum distance of influence. The right map corresponds to the distances to the boundary.

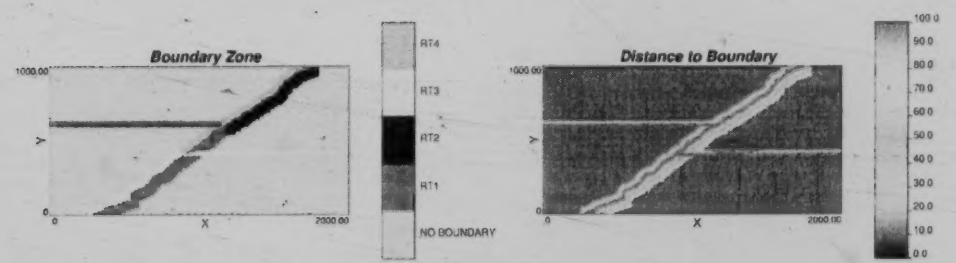

Figure 4.6: Output of the boundmod program for the example of Figure 4.3, considering a neutral arrangement. The left map corresponds to the categorical variable boundary zone. The right map corresponds to the distances to the boundary.

Mean optimization

In order to estimate unsampled locations in a geological model, the mean need to be calculated, variance and covariance model from the available data that involves the optimization of the stationary and non-stationary components. In this section, we will review the implementation of the mean calculation and optimization in a FORTRAN code program called `opt_mean`.

Given a data set where every sample is tagged with the rock type, boundary zone and distance to the boundary, this program calculates the stationary component of the mean for each rock type code and the optimum intercept a_{kp} for each pair: rock type – boundary zone. The parameter a_{kp}, as shown in the previous chapter is the intercept at a distance from the boundary of zero. Since a linear fitting for the non-stationary mean has been chosen for the implementation of this methodology, the optimization of a_{kp} is equivalent to the optimization of the non-stationary mean (Equation 3.8). More complex mean surfaces could be considered by revising the code.

For this, the program first calculates the stationary means for each rock type, calculating the weighted average (if declustering weights are available) over all samples that are within the internal stationary portion, that is, outside any of the boundary zones within the rock type. Then, a boundary zone and parameter a_{kp} is chosen randomly and perturbed within 20% of its original value. Using the modified value of a_{kp}, the objective function corresponding to the sum of the squared differences between the sample value and the mean for all samples is evaluated. The stationary mean of the rock type where the sample belongs plus the non-stationary component, calculated from the distance of the sample to the boundary and the parameter a_{kp} (Equation 3.8) are checked. If the objective function decreases the proposed change to a_{kp} is accepted; if not, the original value of the parameter a_{kp} from the previous iteration is restored. A flow chart with the program structure is shown in Figure 4.7.

The parameter file (Figure 4.8) contains the name of the file with the samples tagged with rock type, boundary zone and distance to the boundary; the column number of the variable, rock type, boundary zone, distance to the boundary and declustering weights; trimming limits to exclude samples, either not assayed or from the upper or lower tails of the distribution. The number of rock types is also required together with the matrix of maximum distances of influence as described for `boundmod` program. The program requires that the user enter the number of iterations. A name for the output file is also required.

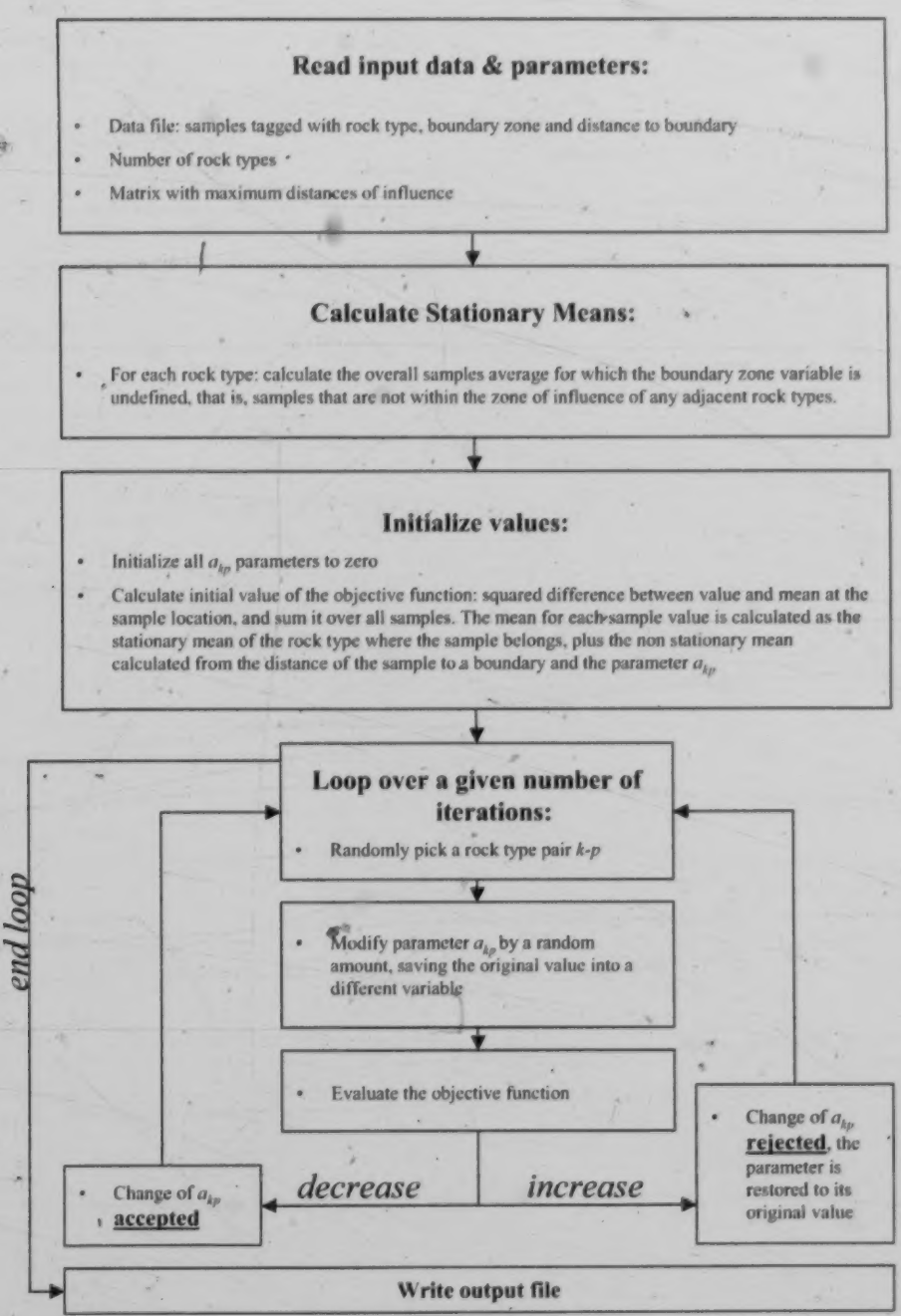
Read input data & parameters:
Data file: samples tagged with rock type, boundary zone and distance to boundary
Number of rock types
Matrix with maximum distances of influence
Calculate Stationary Means:
For each rock type: calculate the overall samples average for which the boundary zone variable is undefined, that is, samples that are not within the zone of influence of any adjacent rock types.
Initialize values:
Initialize all a_{kp} parameters to zero
Calculate initial value of the objective function: squared difference between value and mean at the sample location, and sum it over all samples. The mean for each sample value is calculated as the stationary mean of the rock type where the sample belongs, plus the non stationary mean calculated from the distance of the sample to a boundary and the parameter a_{kp}
Loop over a given number of iterations:
Randomly pick a rock type pair k-p
Modify parameter a_{kp} by a random amount, saving the original value into a different variable
Evaluate the objective function
end loop
decrease
increase
Change of a_{kp} accepted
Change of a_{kp} rejected, the parameter is restored to its original value
Write output file

Figure 4.7: Flow chart of the program `opt_mean`.

```
          Parameters for OPT_MEAN
          ***********************

START OF PARAMETERS
./data/data.dat                    -file with data
1  2  3  4  5                      -columns for var,RT,Boundary RT,Distance,wt
-1.0       1.0e21                  -trimming limits
3                                  -number of rock types
-1.0   50.0   20.0                 -maximum distance of influence of RT 1 into RT j
10.0   -1.0   50.0                 -maximum distance of influence of RT 2 into RT j
20.0   30.0   -1.0                 -maximum distance of influence of RT 3 into RT j
100000                             -number of iterations
output.out                         -output file
```

Figure 4.8: Parameter file for opt_mean program.

Variance optimization

The variance optimization is performed similarly to the mean, by a FORTRAN code program called opt_var. The program uses the analytical expression for the non-stationary means for the different boundary zones and the corresponding stationary means to find the stationary variances and the optimum intercept b_{kp} that define the linear expression of the non-stationary variance for each boundary zone.

The inputs are the same as for the opt_mean program (Figure 4.9); a data file with the samples tagged with their corresponding rock type, boundary zone and distance to the boundary, the number of rock types and the corresponding matrix of maximum distances of influence, plus the output file from the optimization of the means with the stationary means and the intercepts for the non-stationary means.

The residual for each sample is the data value minus the mean. If the sample does not belong to a boundary zone, then the mean is just the stationary mean for the corresponding rock type. In the case where the sample belongs to a boundary zone, the mean will be the stationary mean plus the non-stationary component calculated from Equation 3.8 using the corresponding intercept obtained from the opt_mean program. After all residuals are calculated, the stationary variance is calculated for each rock type using all relevant samples and declustering weights if available. If there are no data for this calculation in one rock type a default value (-999) is assigned.

The optimization procedure (Figure 4.10) has the same structure as the mean optimization program, that is, a boundary zone and parameter b_{kp} are randomly perturbed within an arbitrarily chosen 20% of its previous value. The objective function (Equation 3.9) is then reevaluated, if its value decreases the proposed change is accepted; if not, the original value of the parameter b_{kp} is restored. The

optimization of the non-stationary variance is done through the optimization of b_{kp} (Equation 3.11).

The parameter file (Figure 4.9) has the same entries described in the previous section for the `opt_mean` program, plus the entry for the output file of this program with the stationary means and intercepts for each rock type and boundary zones.

```
          Parameters for OPT_VAR
          **********************

START OF PARAMETERS
  /data/data.dat                  -file with data
1  2  3  4  5                     -columns for var,RT,Boundary RT,Distance,wt
-1.0        1.0e21                -trimming limits
3                                 -number of rock types
  /data/mean.dat                  -file with means by RT and Boundary RT
1  2  3  4                        -columns for RT,Boundary RT,Stationery mean, intercept
-1.0  50.0   20.0                 -maximum distance of influence of RT 1 into RT j
10.0  -1.0   50.0                 -maximum distance of influence of RT 2 into RT j
20.0  30.0   -1.0                 -maximum distance of influence of RT 3 into RT j
100000                            -number of iterations
output.out                        -output file
```

Figure 4.9: Parameter file for `opt_var` program.

Covariance optimization

The covariance optimization corresponds to finds the optimum range of the pseudo stationary model that best fits the experimental spatial correlation of all pairs within a boundary zone. One minus the relative variogram model scaled by the standard deviation of the head and tail values corresponds to the non-stationary covariance. The FORTRAN program that finds the optimum range is called `opt_cov`.

The inputs for this optimization are: (1) a data set tagged with the rock type, boundary zone and distance to the boundary, (2) the stationary components of mean and variance for each rock type plus the intercepts that define the non-stationary components of this statistics for each boundary zone, (3) the number of rock types and matrix of maximum distances of influences, (4) the direct variogram models of the stationary portions of each rock type, and (5) a proposed model (shape, relative nugget effect and anisotropies) for the cross variograms given by the user based on the residuals spatial correlation at the boundary. The anisotropy is specified and fixed through the ratio between the initial ranges input by the modeler. A parameter file is shown in Figure 4.11 as an example.

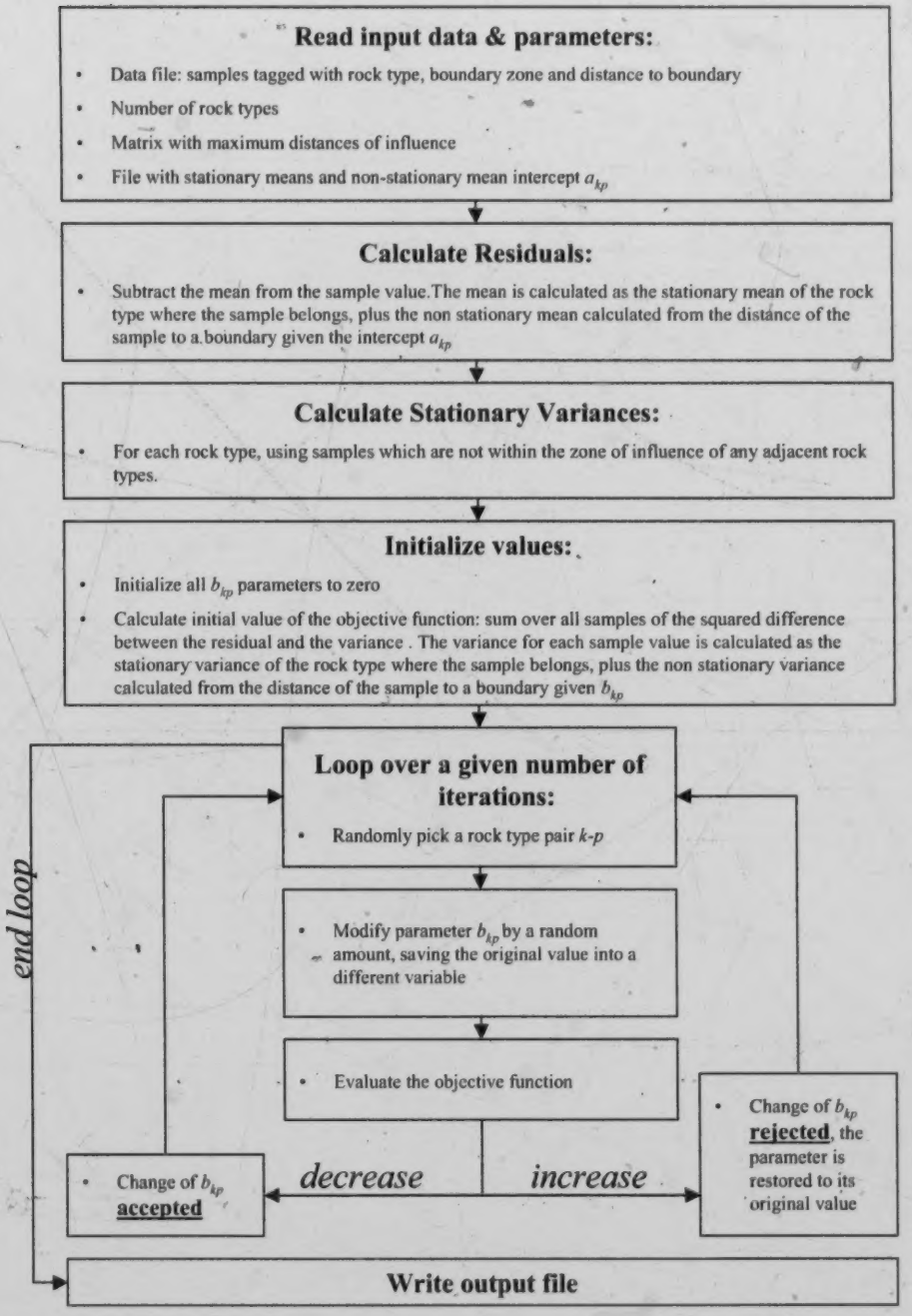
Read input data & parameters:
Data file: samples tagged with rock type, boundary zone and distance to boundary
Number of rock types
Matrix with maximum distances of influence
File with stationary means and non-stationary mean intercept a_{kp}
Calculate Residuals:
Subtract the mean from the sample value.The mean is calculated as the stationary mean of the rock type where the sample belongs, plus the non stationary mean calculated from the distance of the sample to a boundary given the intercept a_{kp}
Calculate Stationary Variances:
For each rock type, using samples which are not within the zone of influence of any adjacent rock types.
Initialize values:
Initialize all b_{kp} parameters to zero
Calculate initial value of the objective function: sum over all samples of the squared difference between the residual and the variance . The variance for each sample value is calculated as the stationary variance of the rock type where the sample belongs, plus the non stationary variance calculated from the distance of the sample to a boundary given b_{kp}
Loop over a given number of iterations:
Randomly pick a rock type pair k-p
Modify parameter b_{kp} by a random amount, saving the original value into a different variable
Evaluate the objective function
decrease
increase
Change of b_{kp} accepted
Change of b_{kp} rejected, the parameter is restored to its original value
end loop
Write output file

Figure 4.10: Flow chart of the program `opt_var`.

The program loops over all possible boundary zones. For each boundary zone, all pairs are found and the experimental covariance is calculated as the multiplication of the head and tail residual values. This experimental covariance is compared against the stationary plus non-stationary covariances derived from the direct variograms models and the proposed cross variogram model to be optimized. To determine the covariance value from a model given a distance, the subroutine of GSLIB `cova3` was used.

Since the modeler fixes the shape, nugget effect and anisotropies, the optimization of the non-stationary covariance model is done through the optimization of the range. The range is iteratively perturbed until the objective function corresponding to the sum over all pair squared differences between the experimental and modeled covariance of each pair, is minimized. The structure of the program is shown in the flow chart of Figure 4.12.

The `opt_cov` output file gives the optimum range for each cross variogram corresponding to the boundary zone defined by a pair of rock type *k* and boundary rock type *p*.

```
                Parameters for OPT_COV
                **********************

START OF PARAMETERS
../data/data.dat                        -file with data
1  2  3  4  5  6  7                     -columns for X,Y,Z,var,RT,Boundary RT,Distance
-1.0       1.0e21                       -trimming limits
100000                                  -number of iterations
output.out                              -output file
3                                       -number of rock types
../data/mean.dat                        -file with means by RT and Boundary RT
1  2  3  4                             -columns for RT,Boundary RT,stationary mean,intercept
../data/var.dat                         -file with variances by RT and Boundary RT
1  2  3  4                             -columns for RT,Boundary RT,stationary variance,intercept
-1.0  50.0   20.0                       -maximum distance of influence of RT 1 into RT j
10.0  -1.0   50.0                       -maximum distance of influence of RT 2 into RT j
20.0  30.0   -1.0                       -maximum distance of influence of RT 3 into RT j
1     1                                 -semivariogram for "RTi" and "RTj"
1   0.1                                 -   nst, nugget effect
1   0.9  0.0   0.0   0.0               -   it,cc,ang1,ang2,ang3
         60.0  60.0  60.0               -   a_hmax, a_hmin, a_vert
1     2                                 -semivariogram for "RTi" and "RTj"
1   0.1                                 -   nst, nugget effect
1   0.4  0.0   0.0   0.0               -   it,cc,ang1,ang2,ang3
         60.0  60.0  60.0               -   a_hmax, a_hmin, a_vert
2     2                                 -semivariogram for "RTi" and "RTj"
1   0.1                                 -   nst, nugget effect
1   0.9  0.0   0.0   0.0               -   it,cc,ang1,ang2,ang3
         60.0  60.0  60.0               -   a_hmax, a_hmin, a_vert
```

Figure 4.11: Parameter file for `opt_cov` program.

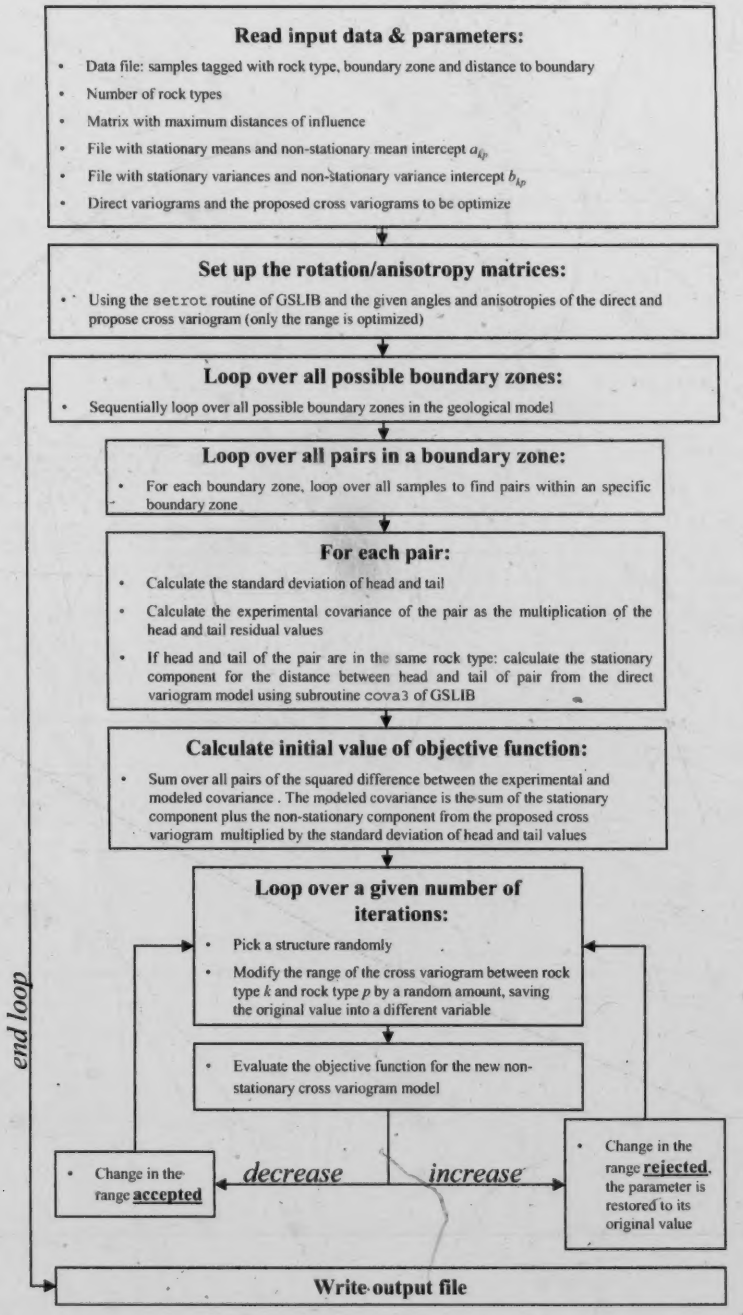
Read input data & parameters:
Data file: samples tagged with rock type, boundary zone and distance to boundary
Number of rock types
Matrix with maximum distances of influence
File with stationary means and non-stationary mean intercept a_{kp}
File with stationary variances and non-stationary variance intercept b_{kp}
Direct variograms and the proposed cross variograms to be optimize
Set up the rotation/anisotropy matrices:
Using the setrot routine of GSLIB and the given angles and anisotropies of the direct and propose cross variogram (only the range is optimized)
Loop over all possible boundary zones:
Sequentially loop over all possible boundary zones in the geological model
Loop over all pairs in a boundary zone:
For each boundary zone, loop over all samples to find pairs within an specific boundary zone
For each pair:
Calculate the standard deviation of head and tail
Calculate the experimental covariance of the pair as the multiplication of the head and tail residual values
If head and tail of the pair are in the same rock type: calculate the stationary component for the distance between head and tail of pair from the direct variogram model using subroutine cova3 of GSLIB
Calculate initial value of objective function:
Sum over all pairs of the squared difference between the experimental and modeled covariance . The modeled covariance is the sum of the stationary component plus the non-stationary component from the proposed cross variogram multiplied by the standard deviation of head and tail values
Loop over a given number of iterations:
Pick a structure randomly
Modify the range of the cross variogram between rock type k and rock type p by a random amount, saving the original value into a different variable
Evaluate the objective function for the new non-stationary cross variogram model
decrease
increase
Change in the range accepted
Change in the range rejected, the parameter is restored to its original value
end loop
Write output file

Figure 4.12: Flow chart of the program `opt_cov`.

Non-stationary cokriging within boundaries

Finally, the program to perform a non-stationary version of simple cokriging in the presence of soft boundaries is the FORTRAN coded program `kt3d_bound`. This program allows to estimate unsampled locations in a geological model where non-stationary boundary zones have been identified, by using a covariance model that takes into account both the stationary and non-stationary components of the conditioning data. It also considers that the mean and variance have non-stationary components at locations within the boundary zones.

The inputs to this program include a data file and rock type model tagged by the corresponding rock type, boundary zone and distance to the boundary of each location, the statistical optimized models for the mean, variance and covariance and the kriging parameters. The output is a GSLIB grid-type file with the estimates and kriging variances for all locations within the grid definition of the model.

The program (Figure 4.13) is a modified version of the GSLIB program `kt3d` (Deutsch and Journel, 1998) that performs simple kriging and/or cross validation using samples from the geological domain of the location to estimate and from adjacent domains. The data-to-data covariance matrix and the data-to-estimate covariance vector (Equation 3.13) are filled according to the spatial configuration of the samples and location being estimated in relation with their location outside or inside a boundary zone and their distances to the boundary (Equation 3.5). The calculated covariance between samples or with the location being estimated includes the non-stationary covariance component corresponding to the optimized model. A detailed scheme of how the covariance matrices and vectors are filled in this type of kriging is presented in Figure 4.14. Once the kriging system is solved, using the `ktsol` routine of GSLIB, the mean: stationary plus non-stationary component, of each sample and the location to be estimated are calculated in order to compute the solution as in Equation 3.12. To calculate the kriging variance (Equation 3.14), the block covariance is calculated for each location to be estimated, as the stationary covariance of the corresponding rock type plus the non-stationary covariance if the block is within a boundary zone.

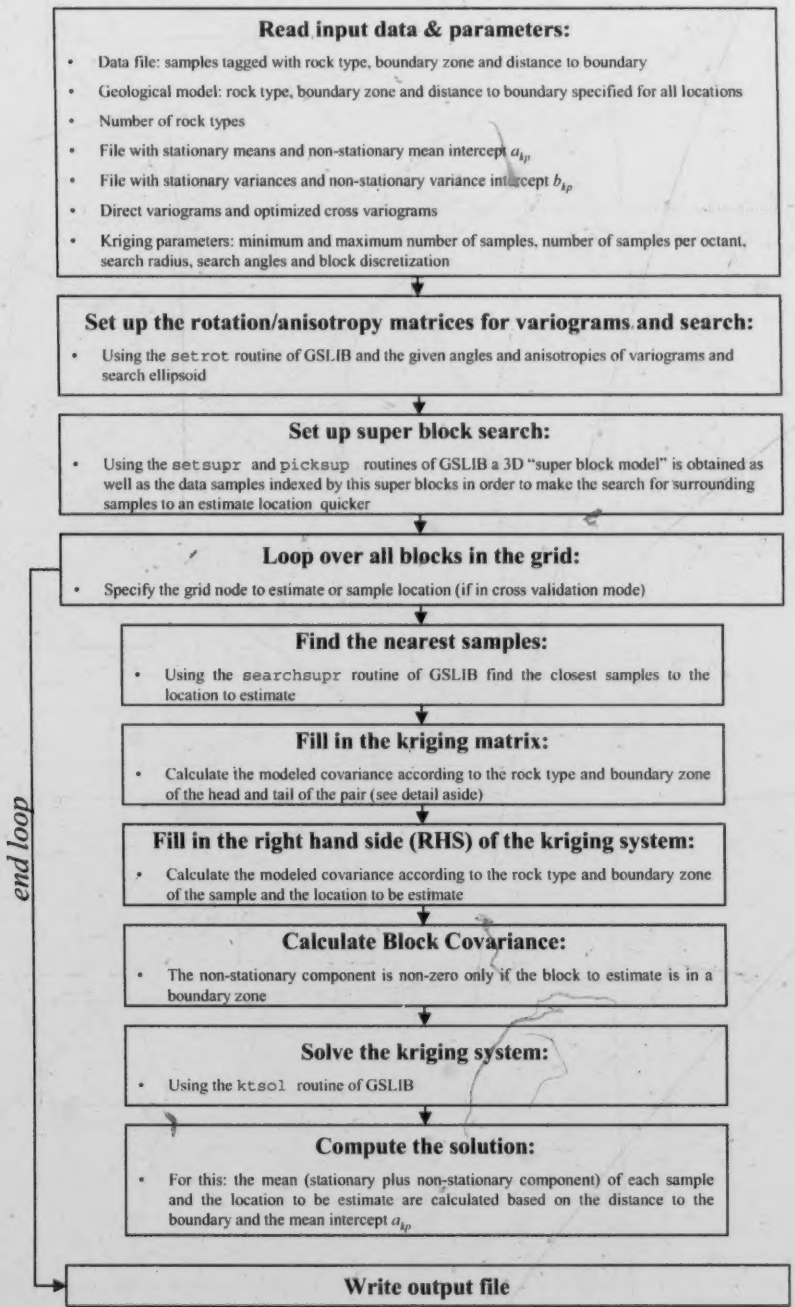
Read input data & parameters:
Data file: samples tagged with rock type, boundary zone and distance to boundary
Geological model: rock type, boundary zone and distance to boundary specified for all locations
Number of rock types
File with stationary means and non-stationary mean intercept a_{kp}
File with stationary variances and non-stationary variance intercept b_{kp}
Direct variograms and optimized cross variograms
Kriging parameters: minimum and maximum number of samples, number of samples per octant, search radius, search angles and block discretization
Set up the rotation/anisotropy matrices for variograms and search:
Using the setrot routine of GSLIB and the given angles and anisotropies of variograms and search ellipsoid
Set up super block search:
Using the setsupr and picksup routines of GSLIB a 3D "super block model" is obtained as well as the data samples indexed by this super blocks in order to make the search for surrounding samples to an estimate location quicker
Loop over all blocks in the grid:
Specify the grid node to estimate or sample location (if in cross validation mode)
Find the nearest samples:
Using the searchsupr routine of GSLIB find the closest samples to the location to estimate
Fill in the kriging matrix:
Calculate the modeled covariance according to the rock type and boundary zone of the head and tail of the pair (see detail aside)
Fill in the right hand side (RHS) of the kriging system:
Calculate the modeled covariance according to the rock type and boundary zone of the sample and the location to be estimate
Calculate Block Covariance:
The non-stationary component is non-zero only if the block to estimate is in a boundary zone
Solve the kriging system:
Using the ktsol routine of GSLIB
Compute the solution:
For this: the mean (stationary plus non-stationary component) of each sample and the location to be estimate are calculated based on the distance to the boundary and the mean intercept a_{kp}
end loop
Write output file

Figure 4.13: Flow chart of the program `kt3d_bound`.

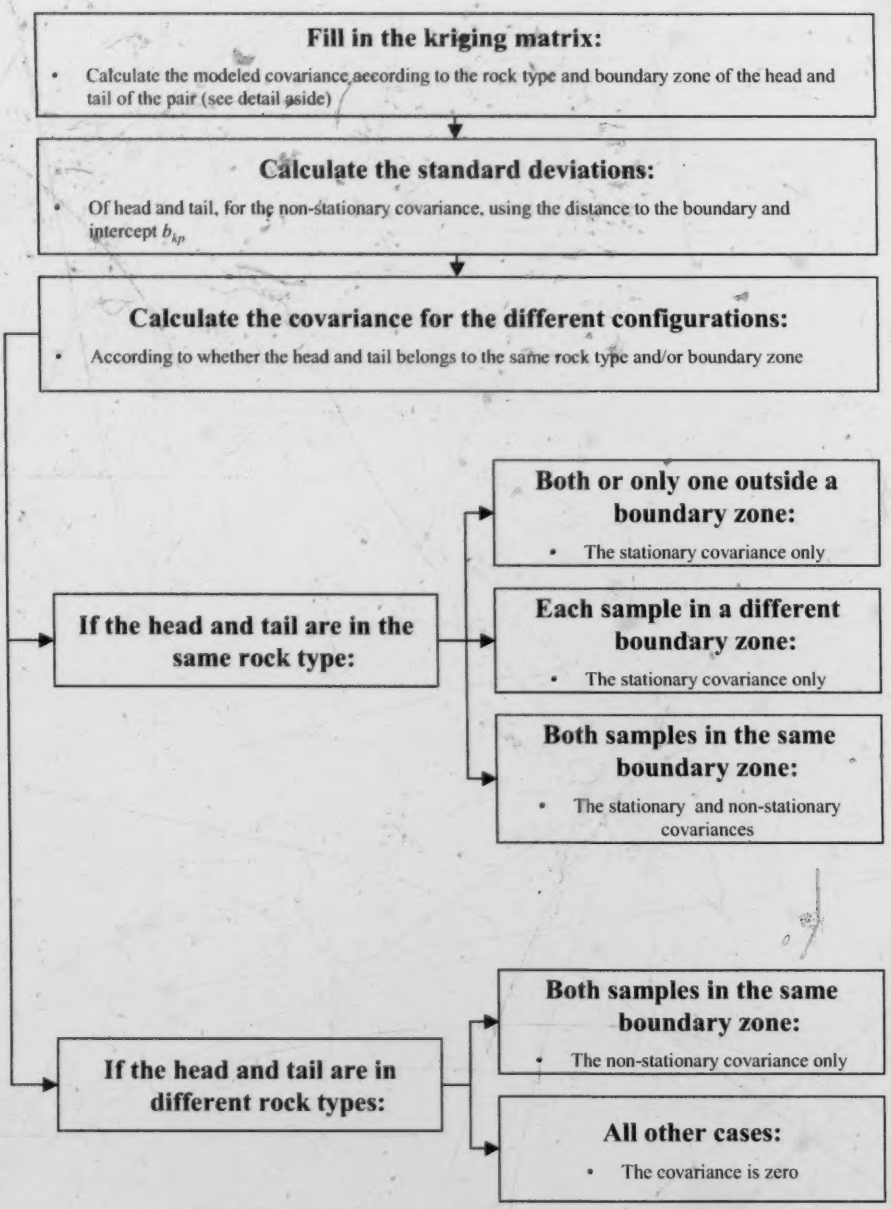
Fill in the kriging matrix:
• Calculate the modeled covariance according to the rock type and boundary zone of the head and tail of the pair (see detail aside)
Calculate the standard deviations:
• Of head and tail, for the non-stationary covariance, using the distance to the boundary and intercept b_{kp}
Calculate the covariance for the different configurations:
• According to whether the head and tail belongs to the same rock type and/or boundary zone
If the head and tail are in the same rock type:
Both or only one outside a boundary zone:
• The stationary covariance only
Each sample in a different boundary zone:
• The stationary covariance only
Both samples in the same boundary zone:
• The stationary and non-stationary covariances
If the head and tail are in different rock types:
Both samples in the same boundary zone:
• The non-stationary covariance only
All other cases:
• The covariance is zero

Figure 4.14: Flow chart for the filling of the kriging matrix in `kt3d_bound` program, the right hand side of the kriging system is filled in a similar way.

The parameter file (Figure 4.15) is divided into four blocks to organize the input of data and parameters. The first block is for data and rock type model input, the second block is to define the grid to be estimated. In the third block, the output files of the mean and variance optimization are specified, the direct variograms of each rock type and the cross variograms between rock types, also the matrix of maximum distances of influences can be entered following the same procedure explained for boundmod program. The fourth block is for the kriging parameters; the program can be run in an estimation or cross validation mode. The ellipsoid search parameters, the minimum, maximum of samples and maximum number per octant are specified in this block too.

```
            Parameters for KT3D_BOUND
            *************************

START OF PARAMETERS:

START OF DATA INPUT:
../data/data.dat                    -file with data
0  1  2  3  4  5  6  7              -columns for DH,X,Y,Z,var,RT,Boundary RT,Distance
-1.0e21      1.0e21                 -trimming limits
../data/rtmodel.out                 -input file with rock type model
1  2  3                            -column with RT,Boundary RT,Distance
3                                   -number of rock types

START OF GRID DEFINITION:
100    0.5      1.0                 -nx,xmn,xsiz
100    0.5      1.0                 -ny,ymn,ysiz
1      0.5      1.0                 -nz,zmn,zsiz

START OF STATISTICAL MODELS:
../data/mean.dat                    -file with means by RT and Boundary RT
1  2  3  4                         -columns for RT,Boundary RT,stationary mean,intercept
../data/var.dat                     -file with variances by RT and Boundary RT
1  2  3  4                         -columns for RT,Boundary RT,stationary variance,intercept
-1.0  50.0   20.0                  -maximum distance of influence of RT 1 into RT j
10.0  -1.0   50.0                  -maximum distance of influence of RT 2 into RT j
20.0  30.0   -1.0                  -maximum distance of influence of RT 3 into RT j
1     1                             -semivariogram for "RTi" and "RTj"
1   0.1                             -   nst, nugget effect
1   0.9  0.0   0.0   0.0           -   it,cc,ang1,ang2,ang3
         60.0  60.0  60.0          -   a_hmax, a_hmin, a_vert
1     2                             -semivariogram for "RTi" and "RTj"
1   0.1                             -   nst, nugget effect
1   0.4  0.0   0.0   0.0           -   it,cc,ang1,ang2,ang3
         60.0  60.0  60.0          -   a_hmax, a_hmin, a_vert
2     2                             -semivariogram for "RTi" and "RTj"
1   0.1                             -   nst, nugget effect
1   0.9  0.0   0.0   0.0           -   it,cc,ang1,ang2,ang3
         60.0  60.0  60.0          -   a_hmax, a_hmin, a_vert

START OF KRIGING PARAMETERS:
0                                   -option: 0=grid, 1=cross, 2=jackknife
xvk.dat                             -file with jackknife data
0  1  2  3  4  5  6  7              -columns for DH,X,Y,Z,var,RT,Boundary RT,Distance
3                                   -debugging level: 0,1,2,3
kt3d.dbg                            -file for debugging output
kt3d.out                            -file for kriged output
4    8                               min, max data for kriging
0                                   -max per octant (0-> not used)
20.0  20.0  20.0                    -maximum search radii
 0.0   0.0   0.0                    -angles for search ellipsoid
```

Figure 4.15: Parameter file for kt3d_bound program.

1-D Example

To illustrate the concepts from Chapter 3 and the programs described above. A small 1-D synthetic example will be used. Three independent unconditional simulations, with different variogram models were used to build the variable $Z(\mathbf{u})$, that will represent the metal grade across a boundary between two rock types. The variogram models were:

$$\mathbf{SGS1} \sim \gamma(\mathbf{h}) = 1.0 \cdot Sph_{(h=15)}(\mathbf{h})$$
$$\mathbf{SGS2} \sim \gamma(\mathbf{h}) = 1.0 \cdot Sph_{(h=30)}(\mathbf{h})$$
$$\mathbf{SGS3} \sim \gamma(\mathbf{h}) = 1.0 \cdot Sph_{(h=10)}(\mathbf{h})$$

The first two consisted of 100 grid points spaced at 1 meter; the third consists of 40 grid points at the same spacing. The small size of the simulated area made the target statistics (m=0 and σ^2=1) difficult to reproduce exactly; one hundred realizations were generated and one realization was chosen with a mean and variance close to the target.

SGS1 and **SGS2** were transformed to a non-standard normal distribution to reflect different average grade and variability across a boundary.

$$\mathbf{SGS1} \sim \mathrm{N}(2.0,1.0)$$
$$\mathbf{SGS2} \sim \mathrm{N}(0.5,0.25)$$

The third simulation was transformed to a non-standard normal distribution but with mean and variance as functions of distance to the boundary. The mean was assumed to follow a linear function like the one described in Equation 3.8, with a symmetric maximum distance of influence of 20 meters and an intercept a_{12}=2.0 while the variance follows a linear function as well, with the same maximum distance of influence and an intercept b_{12}=1.0 (Figure 4.16), that is,

$$m(\mathbf{u}_i) = \frac{(20 - d(\mathbf{u}_i))}{20} \cdot 2 \quad \text{where } 0.5 \le \mathbf{u}_i \le 19.5$$
$$\sigma^2(\mathbf{u}_i) = \frac{(20 - d(\mathbf{u}_i))}{20} \cdot 1 \quad \text{where } 0.5 \le \mathbf{u}_i \le 19.5$$

The transformed boundary values were calculated as follows:

$$Z(\mathbf{u}_i) = \sigma(\mathbf{u}_i) \cdot Y(\mathbf{u}_i) + m(\mathbf{u}_i)$$

where $Y(\mathbf{u}_i)$ is the original simulated value in normal scores. This third variable will be the non-stationary component of the final random variable.

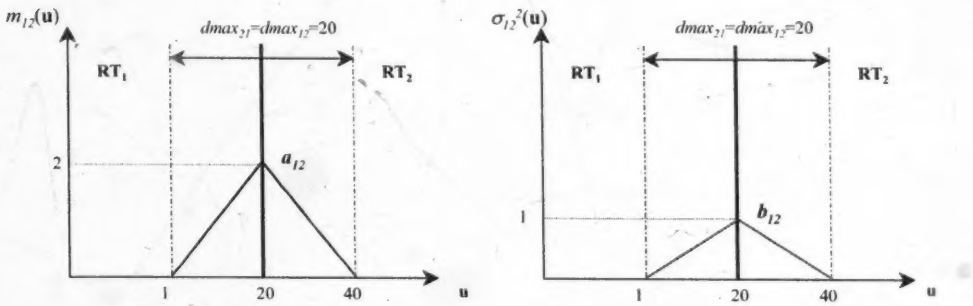

Figure 4.16: Linear functions for the mean and variance used to transform the normal score values of **SGS3**.

The final random variable was obtained by joining the two first simulations to obtain a 1-D array of 200 points and adding the third simulation to the values from locations 80.5 up to 119.5 (Figure 4.17 and 4.18).

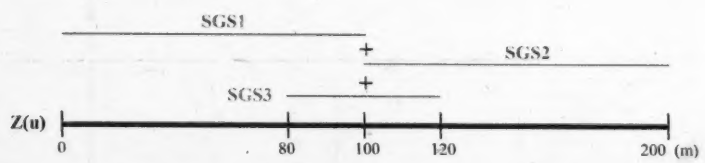

Figure 4.17: Scheme of how the three unconditional simulations are added to obtain $Z(\mathbf{u})$.

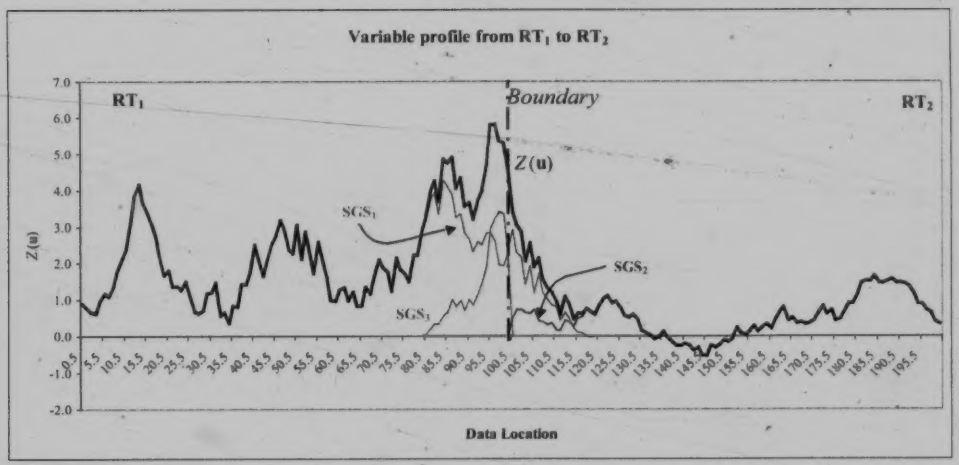

Figure 4.18: Dataset for 1-D example. Random variable profile of metal content against location along the X-coordinate. The colored lines correspond to the underlying variables used to generate this synthetic example.

The first step in this methodology is to infer all stationary and non-stationary statistical parameters for each rock type and boundary zone. To find the stationary and non-stationary component of the mean the program `opt_mean`, was used.

The stationary mean of rock type *1*, calculated with the values from location 0.5 to 79.5, is 1.75, which is lower than the mean of the full **SGS1** (2.0), because the eighty first values are relatively lower than the 20 values closest to the boundary that are not considered in the calculation of the stationary mean. This particular distribution of values is due to ergodic fluctuations in the simulation. The stationary mean of rock type *2* is 0.51.

For this example, after 100,000 iterations, the optimum intercept value a_{21} is 3.22, therefore the mean of $Z(\mathbf{u})$ is,

$$m(\mathbf{u}_i) = \begin{cases} 1.75 & \text{where } 0.5 \le \mathbf{u}_i \le 79.5 \\ 1.75 + \dfrac{(20 - d(\mathbf{u}_i))}{20} \cdot 3.22 & \text{where } 80.5 \le \mathbf{u}_i \le 99.5 \\ 0.51 + \dfrac{(20 - d(\mathbf{u}_i))}{20} \cdot 3.22 & \text{where } 100.5 \le \mathbf{u}_i \le 199.5 \\ 0.51 & \text{where } 120.5 \le \mathbf{u}_i \le 199.5 \end{cases} \quad (4.1)$$

The intercept is slightly higher than the original value of 2.0, used to create this synthetic data, but this is due again to the influence of systematically higher values in $\mathbf{RT_1}$ near the boundary. The discontinuity in the optimum mean profile across the two rock types, after the boundary (Figure 4.19) is a consequence of the difference between the stationary means of $\mathbf{RT_1}$ and $\mathbf{RT_2}$.

The output file of `opt_mean` is shown in Figure 4.20, this file gives the stationary means of rock types *1* and *2* and the intercept of the linear function that represents the non-stationary mean at the boundary zone.

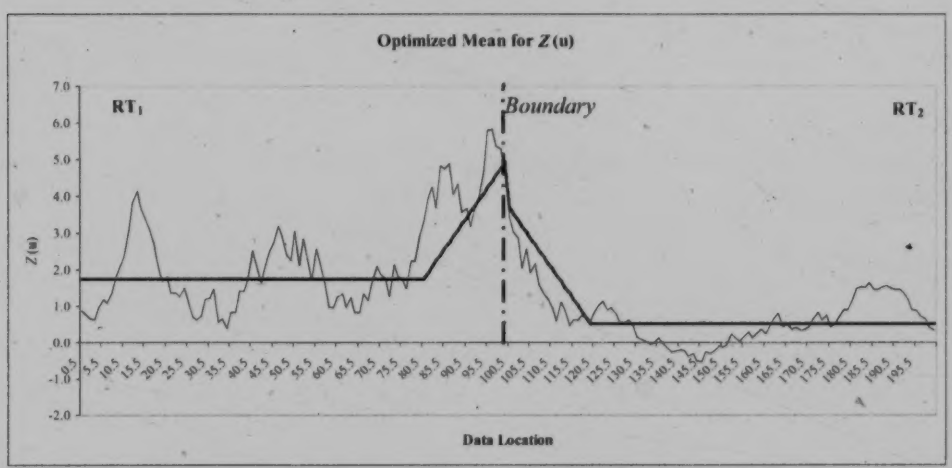

Figure 4.19: Optimized mean obtained for dataset of 1-D example.

To find the stationary and non-stationary component of the variance the program `opt_var` was used. This program uses the output file of `opt_mean` as an input. The objective function in this case is the squared difference between the stationary plus the non-stationary variance and the residuals squared. The residuals are obtained using the already optimized expression for the mean (Equation 4.1). The stationary component of variance for rock type *1* is 0.71 and 0.33 for rock type *2*. The optimum intercept value of b_{21}, in this example is attained at 0.50, thus the total variance of $Z(\mathbf{u})$ (Figure 4.21) is,

$$\sigma^2(\mathbf{u}_i) = \begin{cases} 0.71 & \text{where } 0.5 \le \mathbf{u}_i \le 79.5 \\ 0.71 + \dfrac{(20 - d(\mathbf{u}_i))}{20} \cdot 0.50 & \text{where } 80.5 \le \mathbf{u}_i \le 99.5 \\ 0.33 + \dfrac{(20 - d(\mathbf{u}_i))}{20} \cdot 0.50 & \text{where } 100.5 \le \mathbf{u}_i \le 199.5 \\ 0.33 & \text{where } 120.5 \le \mathbf{u}_i \le 199.5 \end{cases}$$

The optimized intercept is slightly different to the one used to created this synthetic dataset, due to statistical fluctuations from the mean and variance near the boundary zone. The output file of `opt_var` is shown in Figure 4.22.

```
Optimized Means
 4
RT
Boundary_RT
Stationary_Mean
NonStationary_Intercept
    1     2      1.7491          3.2245
    2     1     0.51318          3.2245
```

Figure 4.20: Output file of `opt_mean` program for the 1-D example.

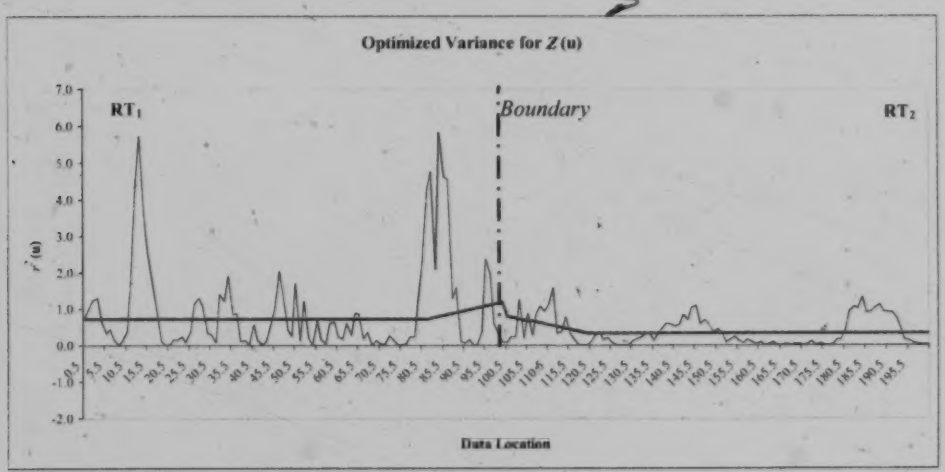

Figure 4.21: Optimized variance obtained for 1-D example.

```
Optimized Variances
 4
RT
Boundary_RT
Stationary_Variance
NonStationary_Intercept
    1    2    0.70569          0.49755
    2    1    0.33372          0.49755
```

Figure 4.22: Output file of `opt_var` program for the 1-D example.

For the estimation we also need the covariance models for both for the stationary and the non-stationary regions in the model. For the stationary regions of $\mathbf{RT}_1$ and $\mathbf{RT}_2$, we will assume that the variogram models are the ones used to generate the underlying unconditional simulations. The FORTRAN code called `opt_cov` is used to find the optimum range of the non-stationary covariance structure that fits the experimental covariance (Figure 4.23) calculated for each pair within the 40 meter zone of influence of the boundary. As explained previously, the modeler must specify the shape and nugget effect of the relative variogram model for the non-stationary zone, while the range is optimized. In this case a spherical isotropic model with a nugget effect of 0.0 was adopted. The optimum range obtained is 6.37 (Figure 4.24), acceptably similar to the range used to build this synthetic dataset given the natural variations we have seen for mean and variance.

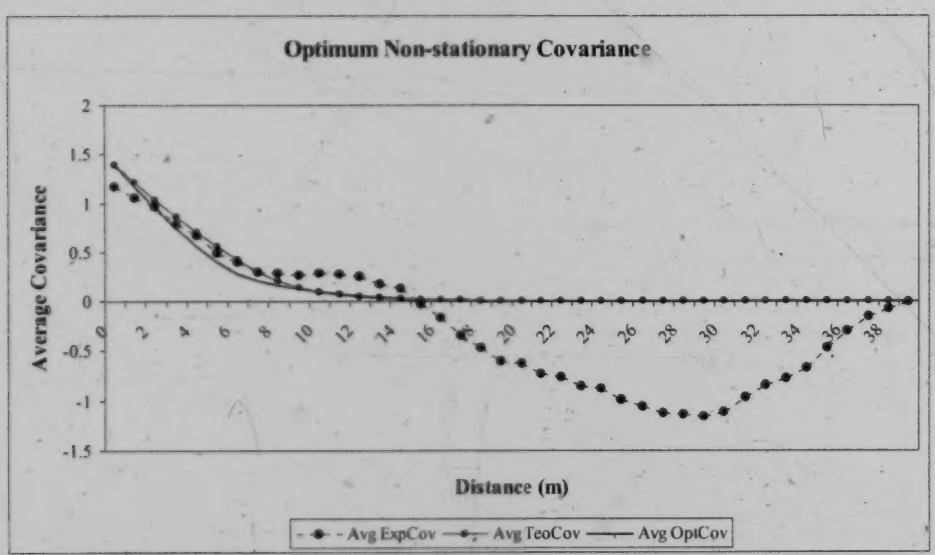

Figure 4.23: Experimental covariance from pairs within the boundary zone (black dash line), optimum non-stationary covariance obtained from `opt_cov` (red solid line) and original covariance of the non-stationary component, **SGS3**, used to build the synthetic dataset $Z(\mathbf{u})$ (blue solid line).

```
Optimized Ranges
 3
RT
Boundary_RT
Optimum_Range
    1    2        6.369907
```

Figure 4.24: Output file of `opt_cov` program for the 1-D example.

The optimization convergence in this example of the mean and variance occur before 250 iterations, while for the covariance, convergence is achieved with less than 50 iterations as shown in Figure 4.25.

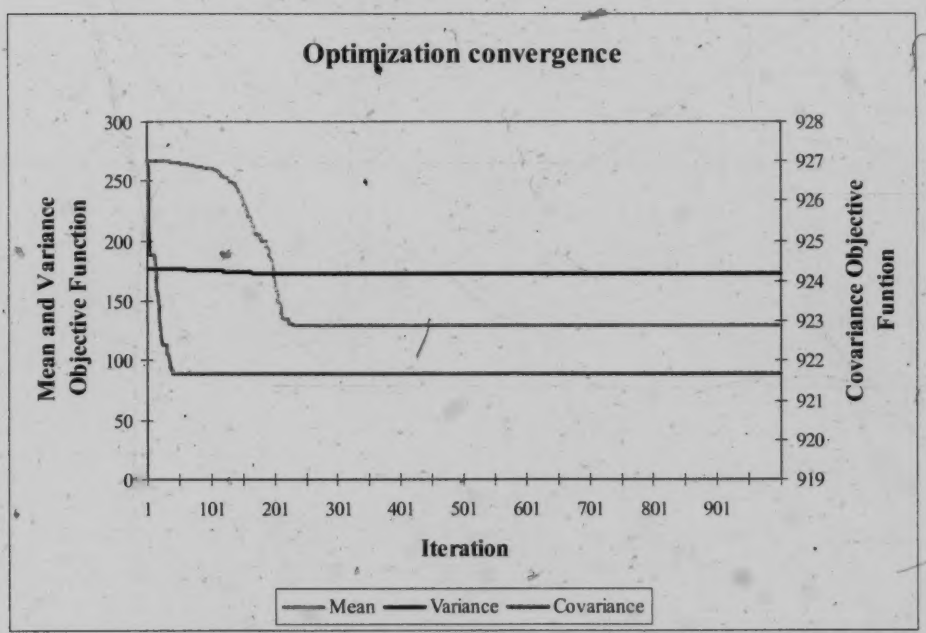

Figure 4.25: Optimization convergence of mean, variance and covariance parameters.

Estimation in the presence of soft boundaries is done using `kt3d_bound` using the mean, variance and covariance previously obtained from optimization. First we will show how estimation is performed with 8 surrounding data (Figure 4.26) at a single location, and then we will review the results of estimation considering a conditioning dataset of one out of four grid nodes from the reference.

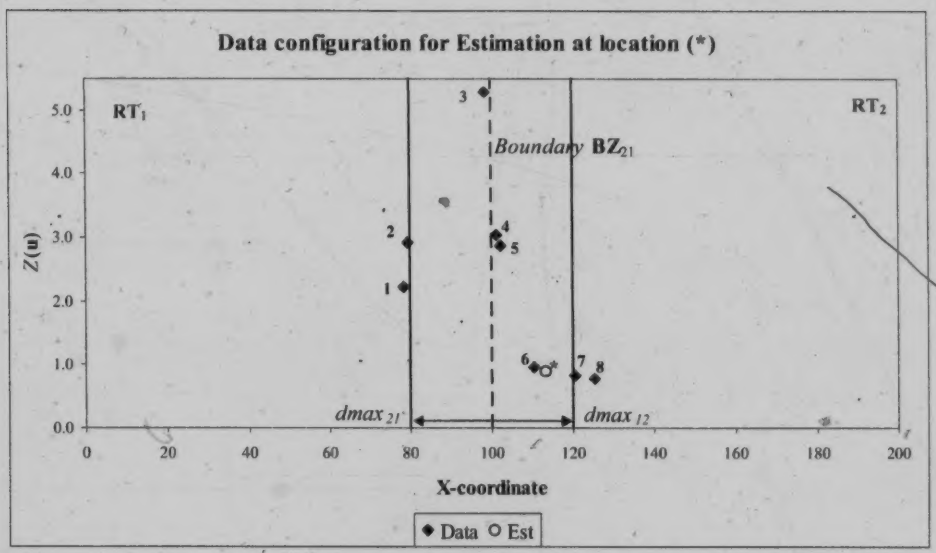

Figure 4.26: Data configuration for the estimation of an unknown location with the 1-D example.

The data covariance matrix in terms of the stationary and non-stationary component, for the example configuration is:

$$
\begin{bmatrix}
Cov(0)^{S}_{Z_1} & & & & & & & \\
Cov^{S}_{Z_1} & Cov(0)^{S}_{Z_1} & & & & & & \\
Cov^{S}_{Z_1}+Cov^{NS}_{Z_{12}} & Cov^{S}_{Z_1}+Cov^{NS}_{Z_{12}} & Cov(0)^{S}_{Z_1}+Cov(0)^{NS}_{Z_{12}} & & & & & \\
0.0 & 0.0 & Cov^{NS}_{Z_{12}} & Cov(0)^{S}_{Z_2}+Cov(0)^{NS}_{Z_{12}} & & & & \\
0.0 & 0.0 & Cov^{NS}_{Z_{12}} & Cov^{S}_{Z_2}+Cov^{NS}_{Z_{12}} & Cov(0)^{S}_{Z_2}+Cov(0)^{NS}_{Z_{12}} & & & \\
0.0 & 0.0 & Cov^{NS}_{Z_{12}} & Cov^{S}_{Z_2}+Cov^{NS}_{Z_{12}} & Cov^{S}_{Z_2}+Cov^{NS}_{Z_{12}} & Cov(0)^{S}_{Z_2}+Cov(0)^{NS}_{Z_{12}} & & \\
0.0 & 0.0 & 0.0 & Cov^{S}_{Z_2}+Cov^{NS}_{Z_{12}} & Cov^{S}_{Z_2}+Cov^{NS}_{Z_{12}} & Cov^{S}_{Z_2}+Cov^{NS}_{Z_{12}} & Cov(0)^{S}_{Z_2} & \\
0.0 & 0.0 & 0.0 & Cov^{S}_{Z_2}+Cov^{NS}_{Z_{12}} & Cov^{S}_{Z_2}+Cov^{NS}_{Z_{12}} & Cov^{S}_{Z_2}+Cov^{NS}_{Z_{12}} & Cov^{S}_{Z_2} & Cov(0)^{S}_{Z_2}
\end{bmatrix}
$$

Similarly, the data-to-estimate covariance vector in this case is:

$$\begin{bmatrix} 0.0 \\ 0.0 \\ Cov_{Z_{12}}^{NS} \\ Cov_{Z_2}^{S} + Cov_{Z_{12}}^{NS} \\ Cov_{Z_2}^{S} + Cov_{Z_{12}}^{NS} \\ Cov_{Z_2}^{S} + Cov_{Z_{12}}^{NS} \\ Cov_{Z_2}^{S} + Cov_{Z_{12}}^{NS} \\ Cov_{Z_2}^{S} + Cov_{Z_{12}}^{NS} \end{bmatrix}$$

Completing the data covariance matrix and the data–to-estimate covariance vector, calculating the stationary and non-stationary component as shown before, the resultant kriging system is:

$$\begin{bmatrix} 1.0 & 0.9 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.9 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 2.149 & 0.382 & 0.227 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.382 & 1.027 & 0.842 & 0.141 & 0.044 & 0.014 \\ 0.0 & 0.0 & 0.227 & 0.842 & 1.003 & 0.152 & 0.052 & 0.019 \\ 0.0 & 0.0 & 0.0 & 0.141 & 0.152 & 0.805 & 0.130 & 0.078 \\ 0.0 & 0.0 & 0.0 & 0.044 & 0.052 & 0.130 & 0.25 & 0.188 \\ 0.0 & 0.0 & 0.0 & 0.014 & 0.019 & 0.078 & 0.188 & 0.25 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \\ \lambda_8 \end{bmatrix} = \begin{bmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 0.108 \\ 0.119 \\ 0.422 \\ 0.164 \\ 0.108 \end{bmatrix}$$

Given that the data covariance matrix is invertible, and calculating the mean at each data location and the mean at the estimate location, as the stationary plus the non-stationary mean, the estimated value is 1.15. The "true" value at this point was 0.89.

The reproduction of the reference values using a conditioning dataset of one out of four grid nodes is fairly good as shown in Figure 4.27 and 4.28. The only problem arises at the edges of the boundary zone, where unusual kriging weights occur leading to discrepancies between the estimate and reference. These weights originate because the covariance of the estimate to data is higher than the covariance of the data to itself due to a non-stationary component in the first one, but not for the data-to-data covariance; in this case the estimate is inside the boundary zone, while the sample is outside. For example, the kriging system for the block at 80.5 meters is,

$$\begin{bmatrix} 1.0 & 0.748 & 0.610 & 0.276 & 0.276 & 0.056 & 0.056 & 0.0 \\ 0.748 & 1.793 & 0.276 & 0.765 & 0.056 & 0.276 & 0.0 & 0.056 \\ 0.610 & 0.276 & 1.0 & 0.056 & 0.610 & 0.0 & 0.276 & 0.0 \\ 0.276 & 0.765 & 0.056 & 1.892 & 0.0 & 0.783 & 0.0 & 0.276 \\ 0.276 & 0.056 & 0.610 & 0.0 & 1.0 & 0.0 & 0.610 & 0.0 \\ 0.056 & 0.276 & 0.0 & 0.783 & 0.0 & 1.992 & 0.0 & 0.802 \\ 0.056 & 0.0 & 0.276 & 0.0 & 0.610 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.056 & 0.0 & 0.276 & 0.0 & 0.802 & 0.0 & 2.091 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 = 1.929 \\ \lambda_2 = -0.171 \\ \lambda_3 = -0.575 \\ \lambda_4 = -0.020 \\ \lambda_5 = -0.017 \\ \lambda_6 = 0.028 \\ \lambda_7 = 0.086 \\ \lambda_8 = -0.003 \end{bmatrix} = \begin{bmatrix} 1.447 \\ 0.967 \\ 0.566 \\ 0.351 \\ 0.208 \\ 0.097 \\ 0.026 \\ 0.0 \end{bmatrix}$$

In this case the kriging weight for the closest data sample is unusually high, although the mismatch between the estimate and the reference is not large. There are other examples for which the differences are more dramatic (see application below).

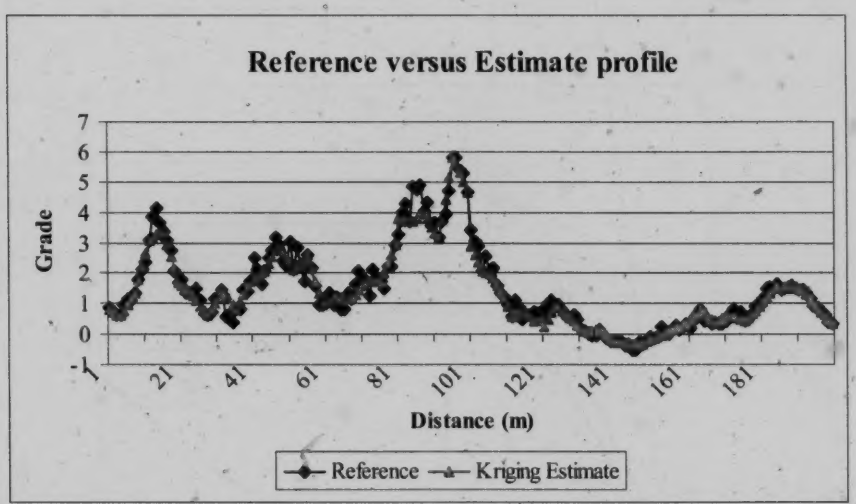

Figure 4.27: Grade reproduction profile along the X-coordinate. Reference values versus kriging estimates.

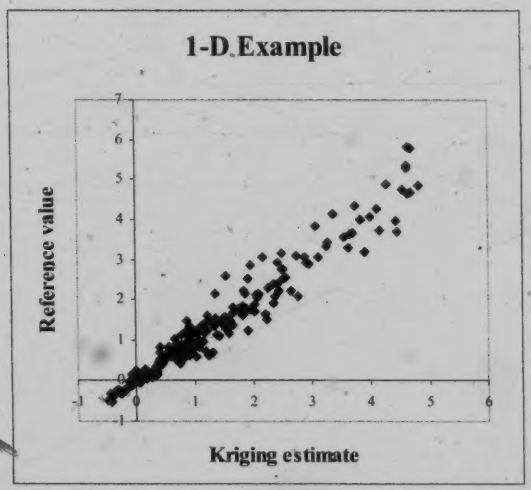

Figure 4.28: Scatter plot reference values versus kriging estimates, for the 1-D example.

For a larger grid, 2000 meters instead of 200, the reproduction of the reference improves (Figure 4.29) as the ergodic fluctuations have less influence in the underlying unconditional simulations use to build the reference and more samples are available to find the stationary and non-stationary components of mean, variance and covariance.

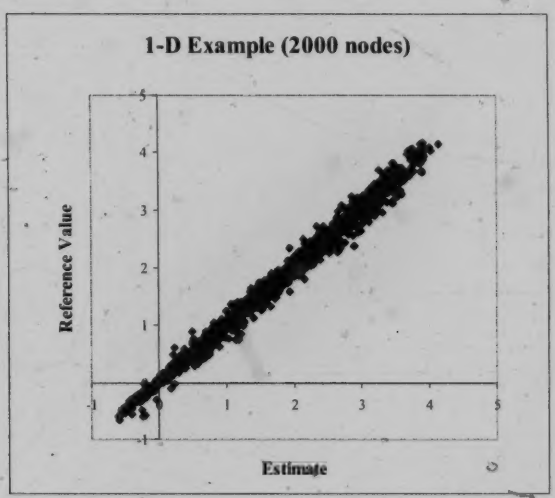

Figure 4.29: Reference values versus kriging estimates considering 2000 nodes instead of 200.

Chapter 5

Application

In this Chapter the application and the steps involved in the non-stationary cokriging in the presence of soft boundaries for a real deposit, will review. The 3-D example was built using the geological model of a porphyry copper deposit from Northern Chile (Figure 5.1), but grades were simulated.

Data generation

An unconditional simulation was generated and transformed to a non-standard normal distribution for each rock type.

$$Y_2 \sim \mathrm{N}(3.0,1.2) \quad \text{with } \gamma_{Y_2}(\mathbf{h}) = 0.2 + 0.8 \cdot Sph_{\begin{pmatrix} h\max=200 \\ h\min=200 \\ vert=100 \end{pmatrix}}(\mathbf{h})$$

$$Y_3 \sim \mathrm{N}(1.0,1.35) \quad \text{with } \gamma_{Y_3}(\mathbf{h}) = 0.1 + 0.9 \cdot Sph_{\begin{pmatrix} h\max=400 \\ h\min=400 \\ vert=100 \end{pmatrix}}(\mathbf{h})$$

$$Y_4 \sim \mathrm{N}(1.5,0.5) \quad \text{with } \gamma_{Y_4}(\mathbf{h}) = 0.1 + 0.9 \cdot Sph_{\begin{pmatrix} h\max=500 \\ h\min=500 \\ vert=100 \end{pmatrix}}(\mathbf{h})$$

$$Y_5 \sim \mathrm{N}(0.25,0.05) \quad \text{with } \gamma_{Y_5}(\mathbf{h}) = 0.2 + 0.8 \cdot Sph_{\begin{pmatrix} h\max=250 \\ h\min=250 \\ vert=100 \end{pmatrix}}(\mathbf{h})$$

All variograms are rotated in 90° azimuth and dip of 35° following the principal anisotropy in the deposit. Rock type 1 corresponds to a leached zone, rock type 2 to a zone of secondary enrichment, rock type 3 and 4 to units of primary mineralisation corresponding to two intrusive events, and rock type 5 to a

peripheral primary mineralisation. The values of rock type 1 were set to a default value (-9) since this unit is of no economical interest.

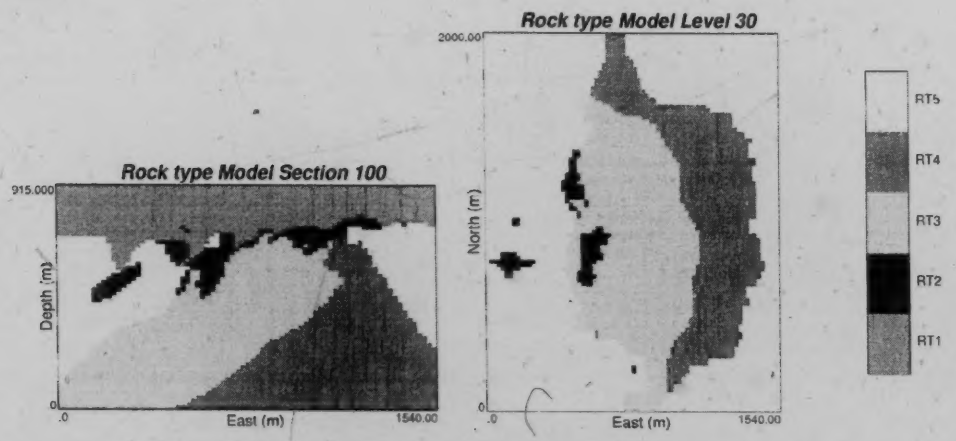

Figure 5.1: Categorical rock type model of a porphyry copper deposit in Northern Chile.

Non-stationary cokriging requires a rock type model with the boundary zone and distance to boundary assigned to each block. The boundmod program requires the matrix of maximum distance of influence and a set of precedence rules as input parameters. Considering the geology of the deposit the matrix of maximum distance of influence (in meters) between rock types was chosen as follows:

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 30 & 30 & 30 \\ 0 & 0 & -1 & 120 & 60 \\ 0 & 0 & 90 & -1 & 90 \\ 0 & 0 & 30 & 30 & -1 \end{bmatrix}$$

The boundary zones defined by the contact between the primary mineralisation units was assumed to be more extended than the ones defined by the secondary mineralisation, since they correspond to a wide fault zone. Also the maximum distance of influence of zero of rock type 3 to 5 into rock type 2 reflects that the secondary mineralisation grades into the primary mineralisation but not the opposite.

The set of precedence rules also reflects the timing of the mineralisation; the influence of rock type 2 is the youngest:

$$\left.\begin{matrix} 4-5 \\ 3-4 \\ 4-3 \end{matrix}\right\} \textit{primary mineralization}$$

$$\left.\begin{matrix} 5-2 \\ 4-2 \\ 3-2 \end{matrix}\right\} \textit{secondary mineralization}$$

For the same section and level shown in Figure 5.1 the boundary zones and distance to boundary are shown in Figure 5.2.

An additional unconditional simulation was generated to build the non-stationary zones around the boundaries:

$$Y_6 \sim \mathrm{N}(0,1) \qquad \text{with } \gamma_{Y_6}(\mathbf{h}) = 0.1 + 0.9 \cdot Sph_{\begin{pmatrix} h\max=400 \\ h\min=400 \\ vert=50 \end{pmatrix}}(\mathbf{h})$$

The variogram of this variable is also rotated in 90° azimuth and dip of 35°.

Y_6 was transformed to a non-stationary variable (Figure 5.3) using the following expression:

$$Y_6^*(\mathbf{u}_i) = \left[\frac{(dmax_{kp} - d(\mathbf{u}_i))}{dmax_{kp}} \cdot b_{kp}\right]^{1/2} \cdot Y_6(\mathbf{u}_i) + \left[\frac{(dmax_{kp} - d(\mathbf{u}_i))}{dmax_{kp}} \cdot a_{kp}\right]$$

where the distance to the boundary as well as the boundary zone were obtained from the rock type model described above. For each boundary zone, a mean and variance intercept were chosen trying to reproduce the real trends in the deposit:

Boundary Zone (RT-BRT)	Intercept a_{kp}	Intercept b_{kp}
3-2	5.0	2.0
4-2	5.0	2.0
5-2	5.0	2.0
3-4 (and 4-3)	3.0	1.0
3-5 (and 5-3)	0.5	0.5
4-5 (and 5-4)	1.0	1.0

Outside the boundary zone this variable was set to zero, that way, it can be added to the grid file with the simulated values Y_2 to Y_5 to obtain the reference (Figure 5.4).

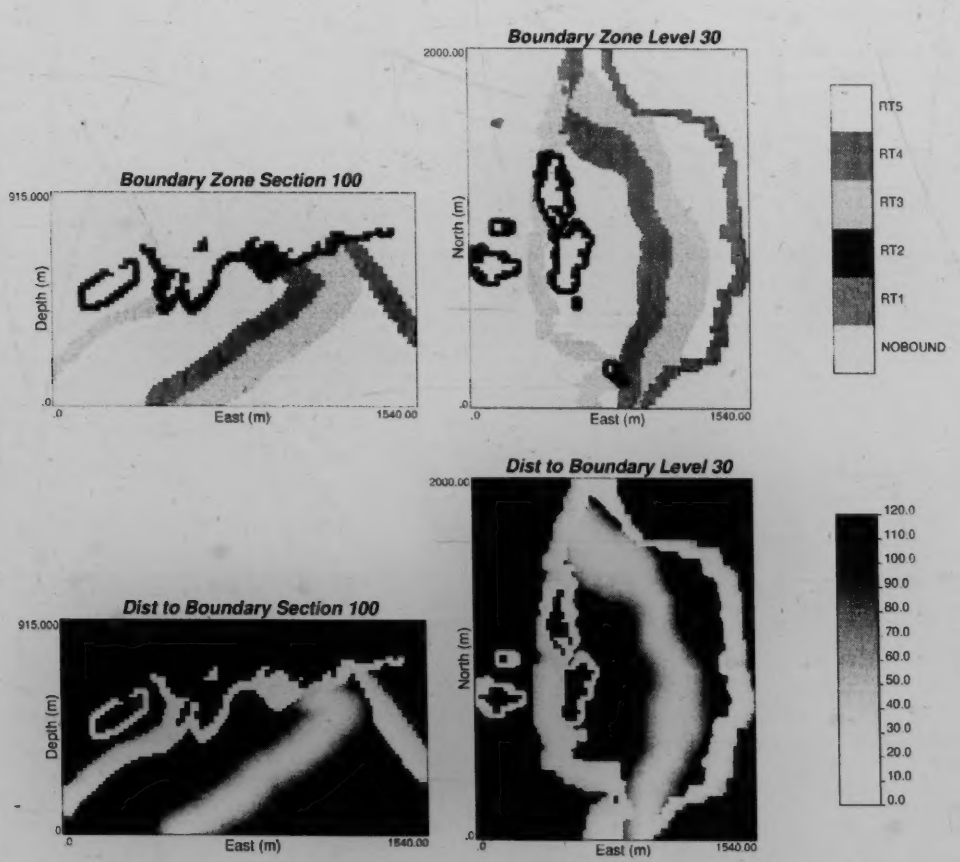

Figure 5.2: Boundary zone (top) and distance to boundary (bottom) maps for the same section and level of the rock type model of Figure 5.1. The default code (NO BOUND) is given to locations beyond the corresponding maximum distance of influence. Distances beyond the maximum distance of influence are set to the default value 9999, in red.

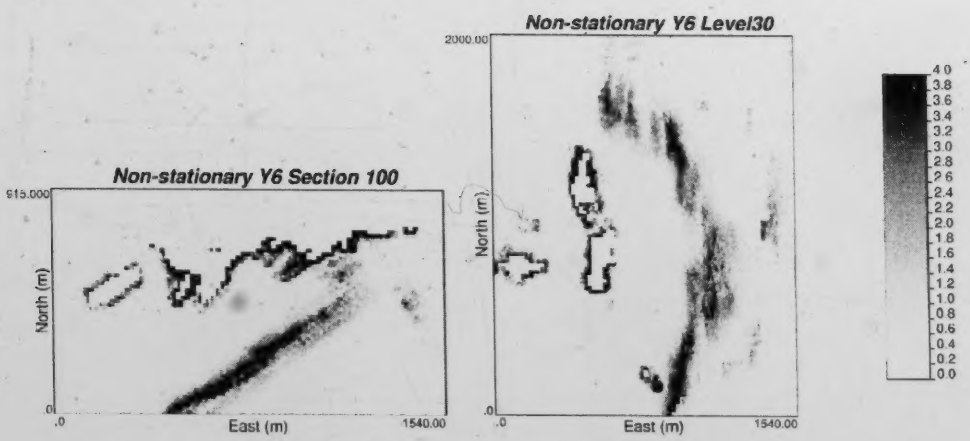

Figure 5.3: Transformed variable $Y_6(\mathbf{u})$, blocks outside a boundary zone were set to zero.

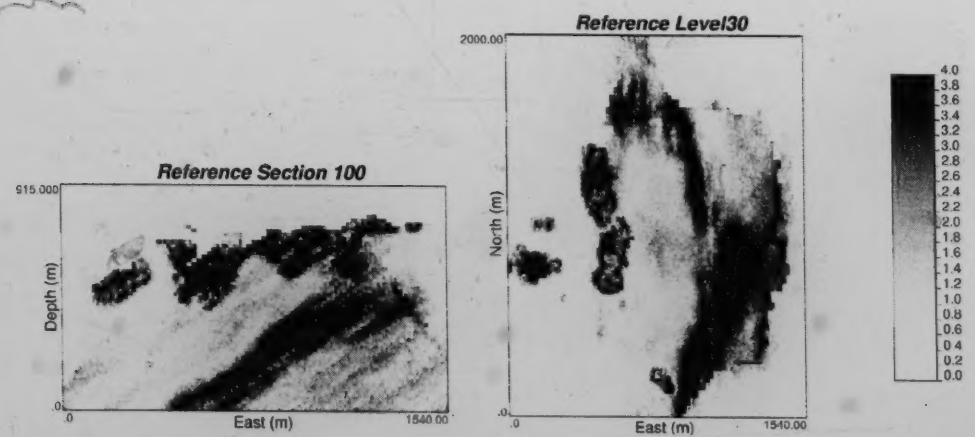

Figure 5.4: Section and Level maps of the reference distribution. Values from rock type 1 were assigned a default value of –9 since this unit is of no economic interest.

Since variables Y_2 to Y_5 were transformed to non-standard normal distributions some negative values occurred. All negative grades (4.3%) were set to zero. Finally the reference grid file was sampled on a 100x100x1 grid to obtain the conditioning data for kriging.

Estimation

Prior to kriging, we need the optimum mean, variance and covariance range for the non-stationary boundaries from the conditioning data. The `opt_mean` output (Figure 5.5) shows that the stationary means of each rock type and the *a* intercepts are well reproduce for each boundary zone compared with the reference (Table 5.1). The *b* intercepts corresponding to the non-stationary variance are acceptably well reproduced for all boundaries (Table 5.1); Figure 5.6 shows the output of `opt_var` program. As already seen in the 1-D example the differences with target parameters increase for the non-stationary variance intercepts and then for the covariance range, since at each optimization step the parameter becomes more sensitive to statistical fluctuations in the previously optimized parameters.

```
Optimized Means
 4
RT
Boundary_RT
Stationary_Mean
NonStationary_Intercept
     1     2    -9.0000          0.0000
     1     3    -9.0000          0.0000
     1     4    -9.0000          0.0000
     1     5    -9.0000          0.0000
     2     1     3.1180          0.0000
     2     3     3.1180          5.0883
     2     4     3.1180          5.7429
     2     5     3.1180          4.9228
     3     1    0.91914          0.0000
     3     2    0.91914          5.0883
     3     4    0.91914          2.9431
     3     5    0.91914         0.60057
     4     1     1.6637          0.0000
     4     2     1.6637          5.7429
     4     3     1.6637          2.9431
     4     5     1.6637          1.1153
     5     1    0.27290          0.0000
     5     2    0.27290          4.9228
     5     3    0.27290         0.60057
     5     4    0.27290          1.1153
```

Figure 5.5: Output file of `opt_mean` program.

In order to find the optimum range of the non-stationary component for each boundary zone, we need: the variogram models of the stationary portions of each

rock type and the shape and nugget effect of the relative standardized variogram. The relative standardized variogram corresponds to the stationary shape that will be scaled by the non-stationary standard deviation at each location.

```
Optimized Variances
 4
RT
Boundary_RT
Stationary_Variance
NonStationary_Intercept
    1     2      0.0000          0.0000
    1     3      0.0000          0.0000
    1     4      0.0000          0.0000
    1     5      0.0000          0.0000
    2     1     0.99811          0.0000
    2     3     0.99811          3.0120
    2     4     0.99811          2.3062
    2     5     0.99811          2.5425
    3     1     0.27125          0.0000
    3     2     0.27125          3.0120
    3     4     0.27125         0.87515
    3     5     0.27125         0.18306
    4     1     0.52029          0.0000
    4     2     0.52029          2.3062
    4     3     0.52029         0.87515
    4     5     0.52029         0.56958
    5     1   0.43380E-01        0.0000
    5     2   0.43380E-01        2.5425
    5     3   0.43380E-01       0.18306
    5     4   0.43380E-01       0.56958
```

Figure 5.6: Output file of `opt_var` program.

Boundary Zone (RT-BRT)	Reference Intercept a_{kp}	Optimum Intercept a_{kp}	Reference Intercept b_{kp}	Optimum Intercept b_{kp}
3-2	5.0	5.09	2.0	3.01
4-2	5.0	5.74	2.0	2.31
5-2	5.0	4.92	2.0	2.54
3-4 (and 4-3)	3.0	2.94	1.0	0.88
3-5 (and 5-3)	0.5	0.60	0.5	0.18
4-5 (and 5-4)	1.0	1.12	1.0	0.57

Table 5.1: Comparison between the reference and optimum intercepts a_{kp} and b_{kp}.

The samples that belong to the stationary portion of each rock type were selected to calculate and model the variograms in the three principal directions. The stationary variance of each rock type was set as the sill. The models are close to the original variograms used in the unconditional simulations except for the minimum horizontal range of rock type 4 that was calculated quite high (Table 5.2).

Rock type	Structure Type	Sill Contribution	Range Hmax	Range Hmin	Range Vert
2	Nugget Effect	0.2503			
	Spherical	0.7477	250.0	250.0	119.0
3	Nugget Effect	0.0300			
	Spherical	0.2400	531.0	373.0	200.0
4	Nugget Effect	0.0476			
	Spherical	0.4724	498.0	627.0	152.0
5	Nugget Effect	0.0092			
	Spherical	0.0308	267.0	248.0	136.0

Table 5.2: Variogram models for the stationary regions of rock type 2 to 5.

The shape, anisotropies and nugget effect of the relative standardized variogram of the non-stationary boundaries were chosen as the same as the variogram used to generate the Y_6 variable. Normally the modeler will have to pick these values based on the geology of the deposit. If that were the case here, the parameters would be similar to the chosen ones, the non-stationary zone is strongly controlled by the fault system that coincides with the major anisotropy, most of the spatial correlation is explained by spherical models, only the nugget effect is difficult to assess.

The optimization of the covariance ranges gives reasonable results (Figure 5.7) for all boundary zones except for the boundary between rock type 2 and rock type 4, which seems a little too high. This result is likely due to the influence of the minimum horizontal range of rock type 4.

```
Optimized Ranges
 3
RT
Boundary_RT
Optimum_Range
     2     3     326.957855
     2     4     660.023438
     2     5     465.359070
     3     4     425.019592
     3     5     335.151306
     4     5     576.773804
```

Figure 5.7: Output file of `opt_cov` program.

In this example the optimization convergence of the mean and variance occurs around 2800 iterations. This is more than the 1-D example of the previous chapter (Figure 5.8), but consistent with the larger number of data. The covariance range convergence occurs around 30 iterations for all boundary zones (Figure 5.9). Considering the number of pairs involved in each calculation this is surprisingly similar to the convergence time of the same parameter in the 1-D example.

The kriging parameters used for the estimation include: a minimum and maximum of 4 samples, no octant search and an isotropic search radius of 650 meters.

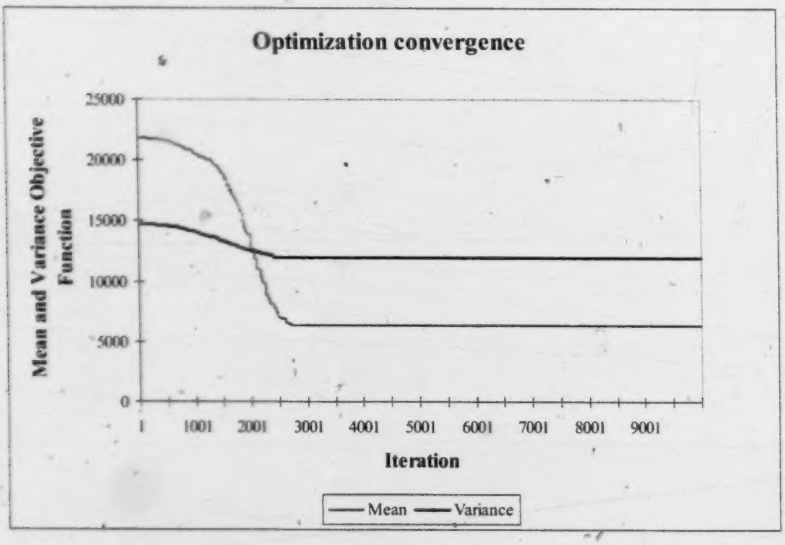

Figure 5.8: Optimization convergence of mean and variance in terms of the objective function value.

In the first run of `kt3d_bound` a significant number of estimates were unusually high or low due to unusual kriging weights. These blocks were concentrated at the edges of the boundary zones. They represent a non-physical covariance model that must be reviewed as part of the future work. Two modifications to the kriging system were made to make the kriging matrix stable for the purpose of this application. A modification is required when the covariance of the estimate to data is higher than the covariance of the data to itself. This occurs because the estimate is inside a boundary zone and therefore has a non-stationary component that is added to the stationary covariance model, while the data is outside. To fix this, the diagonal term corresponding to the covariance of the sample to itself was set to 1.1 multiplied by the corresponding element in the estimate-to-data vector of the kriging system. When the estimation variance is calculated to be negative, all the diagonal terms of the data-to-data matrix were replaced by the maximum value of all elements in the kriging system. Additional to the modifications to the program a relatively small maximum number of samples were used to estimate a block. The origin of these non-physical results needs to be reviewed in the future. These relatively rare

problems in kriging will become more important for the implementation of this technique in simulation where the correct estimate and estimation variance are essential for the reproduction of the conditioning data and its variability.

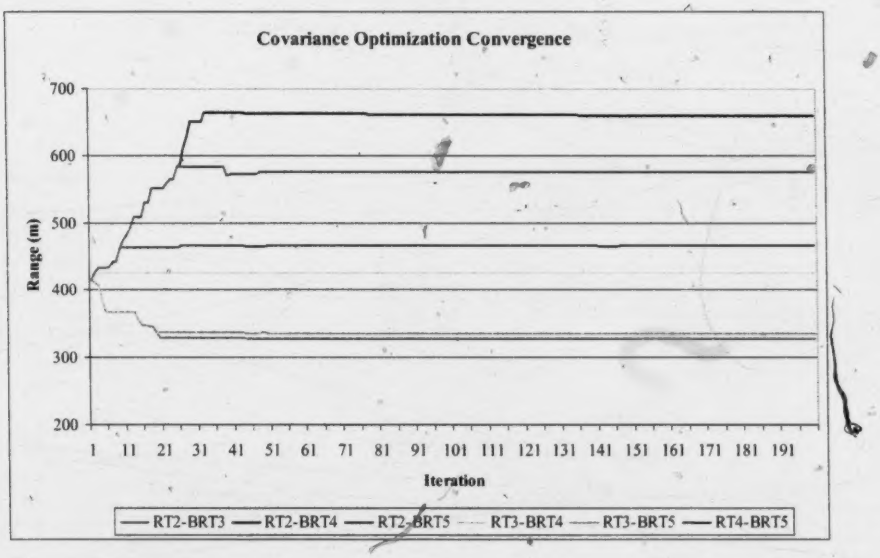

Figure 5.9: Covariance optimization convergence in terms of the range for the 6 boundary zones.

The correlation between the estimates and the reference or 'true' value is around 0.8 for each boundary zone (Figure 5.10). The stationary portions of rock type 2 and 5 show more conditional bias than rock types 3 and 4; the correlation between the estimate and reference ranges from 0.52 up to 0.8 (Figure 5.11). The mean of the reference in the stationary portions of each rock type is reproduced almost exactly by the kriging estimation (Table 5.3). The variance of the estimate is lower than the reference, as expected since kriging has a smoothing effect. The non-stationary behavior of the mean is also very well reproduced by the proposed non-stationary cokriging as shown in Figure 5.12. Although the variance of the estimates in the boundary zone is lower than the reference, the increasing trend toward the boundary is well reproduced (Figure 5.13).

Validation of the model and parameters was done running `kt3d_bound` in the cross validation mode removing the entire drill hole to which the sample belongs. In this mode, the correlation between the estimate and the true value is 0.93. The results show that the resultant model is accurate and precise (Figure 5.14), the

distribution of the error (estimate minus true) is symmetric and centered in zero and has a relatively small standard deviation.

The proposed methodology can be compared with ordinary kriging assuming a traditional soft boundary approach. All data will be combined for a single variogram model and kriging. The proposed methodology significantly outperforms this traditional ordinary kriging approach see Figure 5.15. The correlation coefficient for the proposed methodology is 0.9 versus 0.74 for ordinary kriging, which translates into a 48% improvement in how data is used.

Rock type (stationary region)	Reference		Kriging	
	Mean	Variance	Mean	Variance
2	3.13	1.03	3.14	0.56
3	0.92	0.52	0.91	0.39
4	1.66	0.71	1.66	0.57
5	0.28	0.21	0.27	0.14

Table 5.3: Comparison of the mean and variance of the each rock type (stationary regions) between the reference training image and the kriging results.

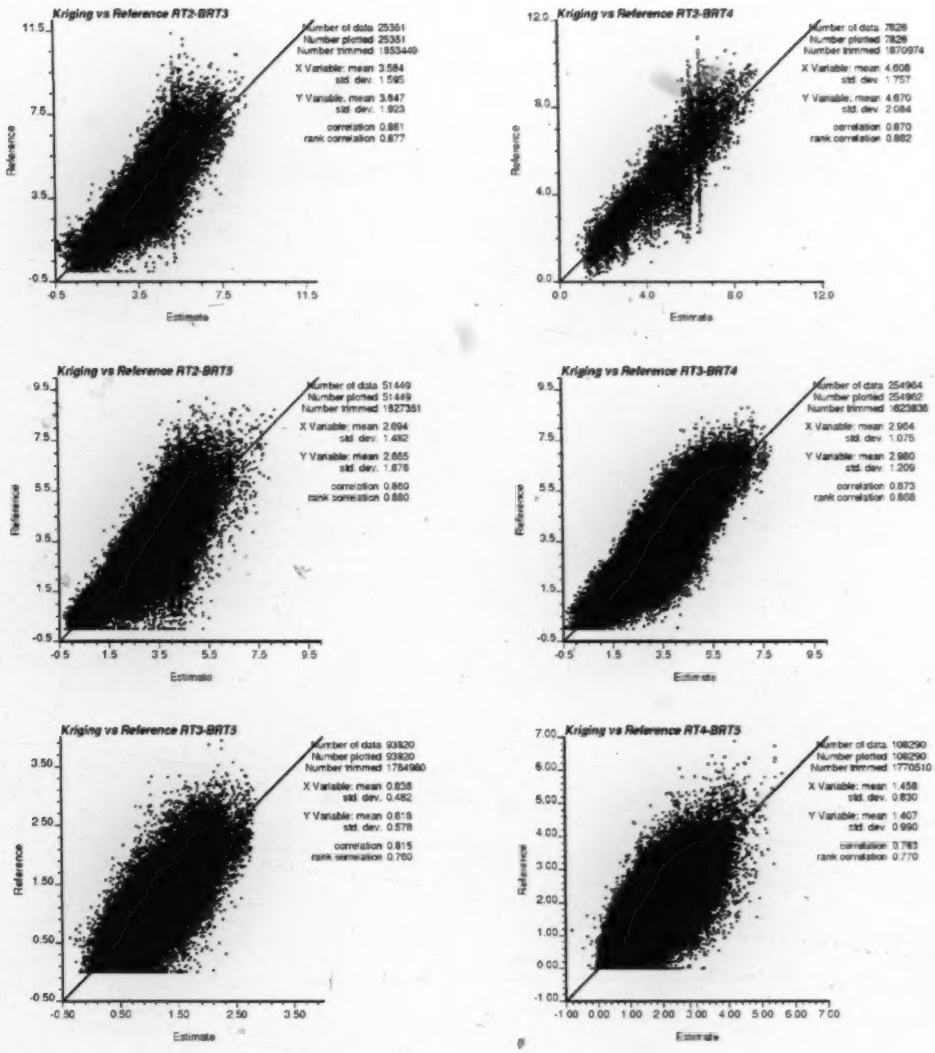

Figure 5.10: Scatter plot of the reference versus the estimate for each boundary zone.

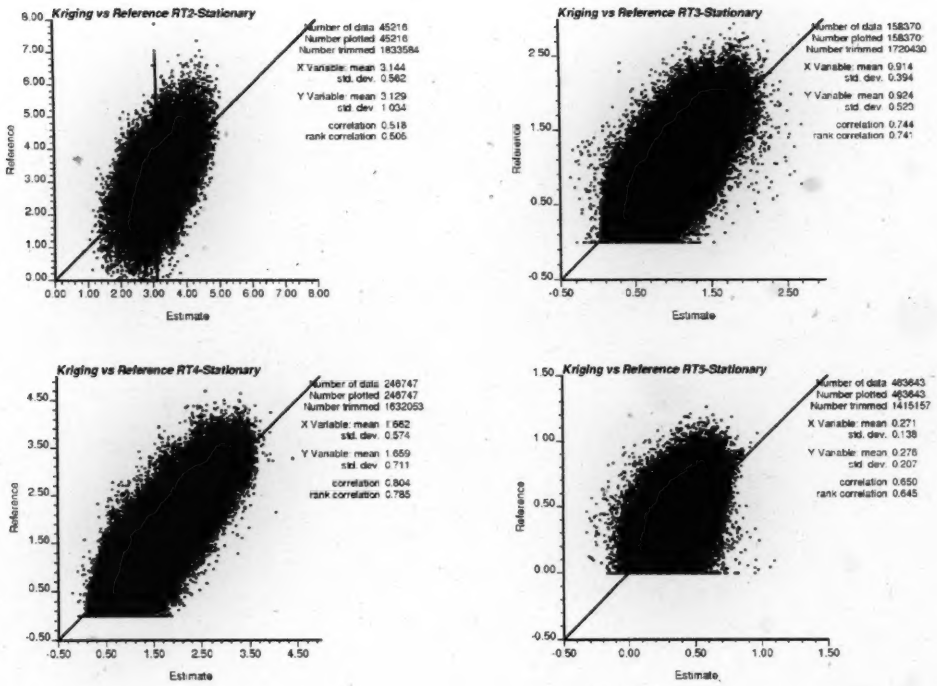

Figure 5.11: Scatter plot of the reference versus the estimate within the stationary portions of each rock type.

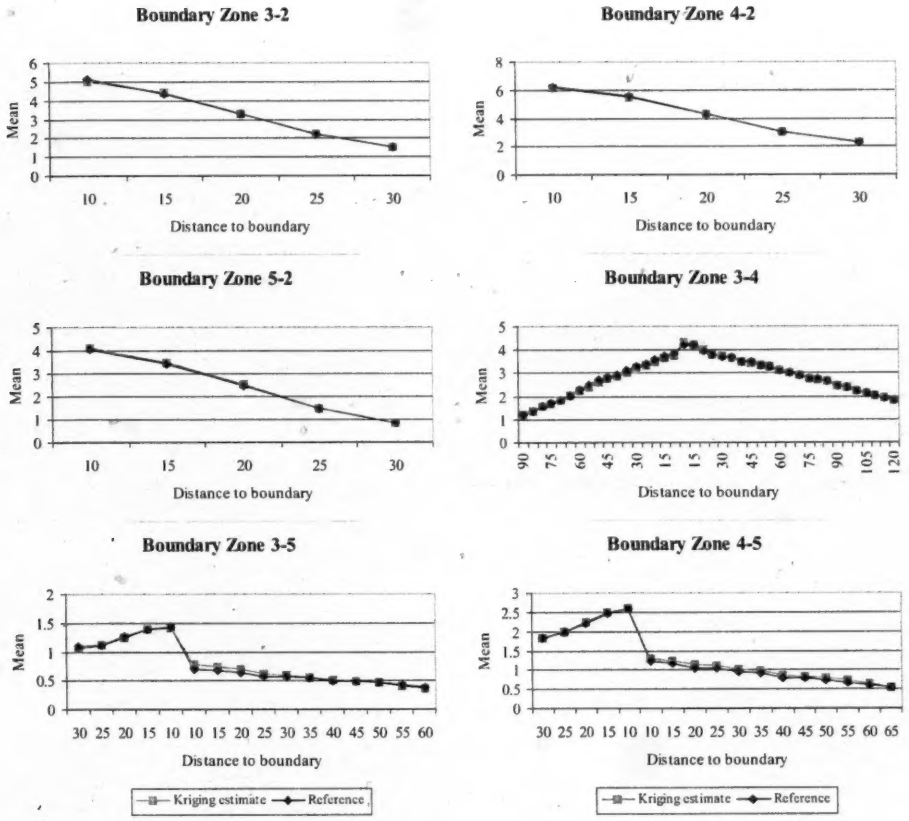

Figure 5.12: Mean at the non-stationary boundary zone. A 5 meters interval of the distance to the boundary was chosen to calculate the mean of the estimate value of all grid nodes.

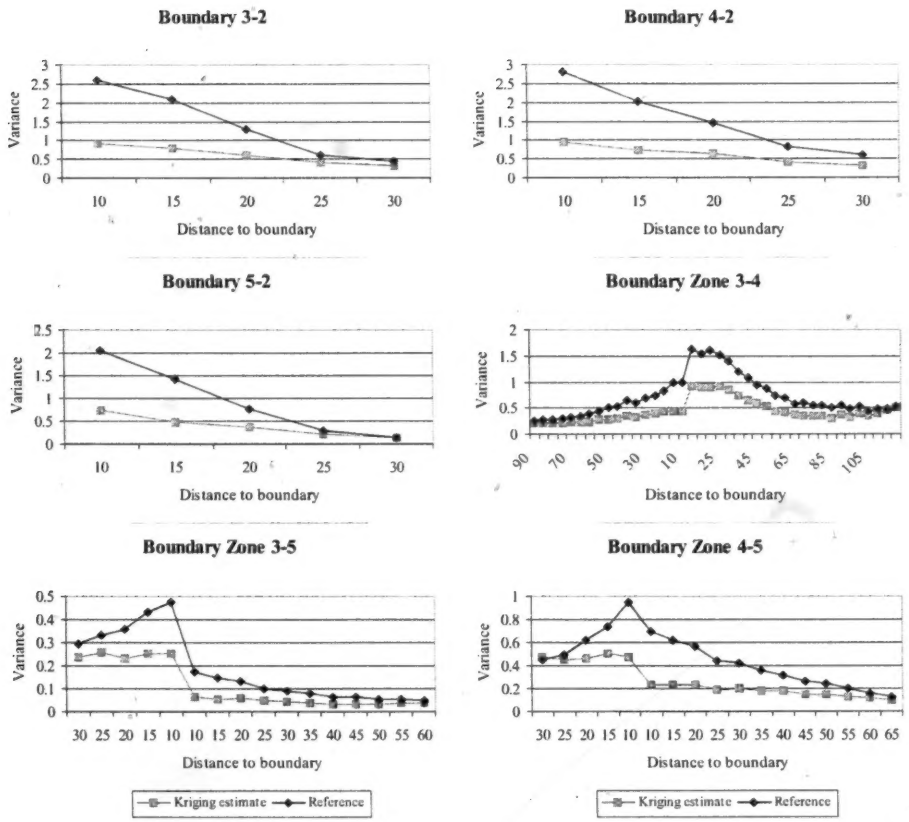

Figure 5.13: Variance at the non-stationary boundary zone. A 5 meters interval of the distance to the boundary was chosen to calculate the variance of the estimate value of all grid nodes.

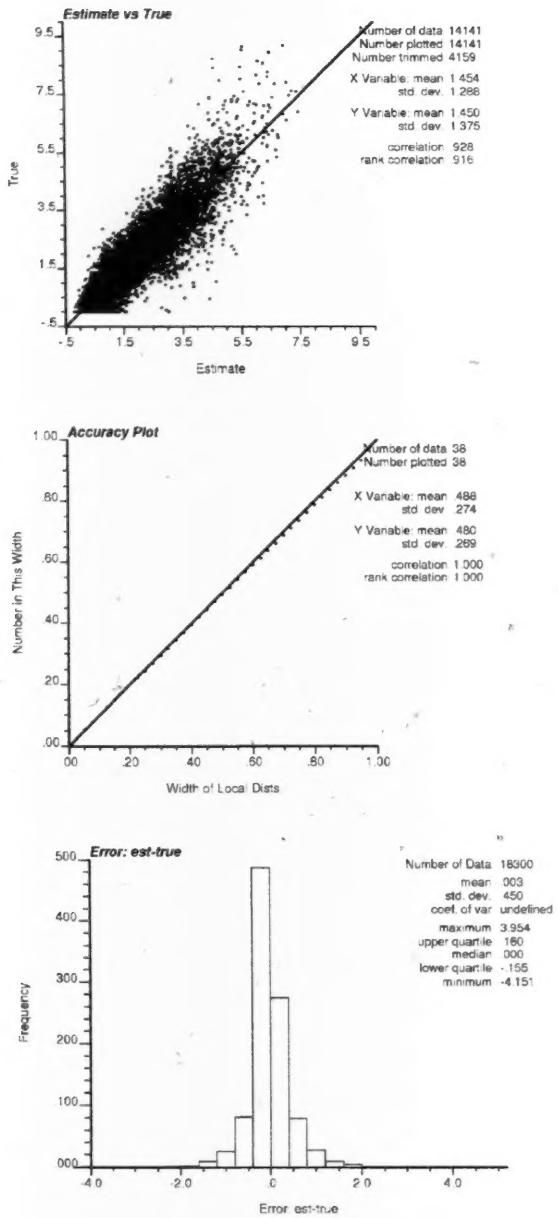

Figure 5.14: Cross validation results. Scatter plot of the estimate versus the true; the correlation is 0.92. The accuracy plot shows that the model is accurate and precise for the chosen parameters. The error (true-estimated) distribution is symmetric and centered in zero and with a relatively low standard deviation.

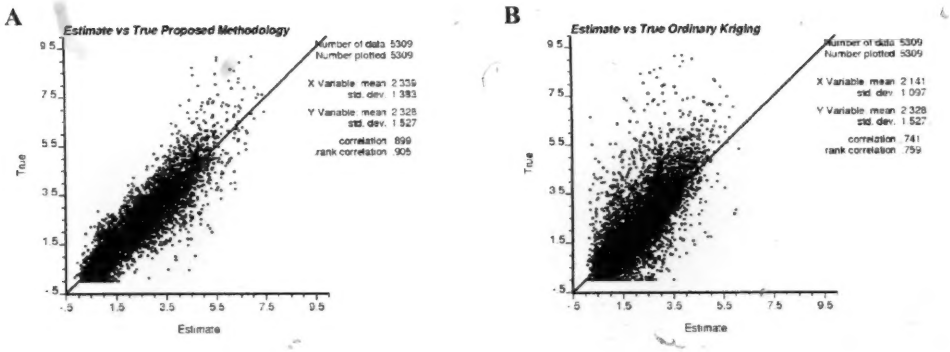

Figure 5.15: Cross validation results comparison at the boundary zone. The proposed methodology (A) has a correlation coefficient of 0.90 versus 0.74 for traditional ordinary kriging approach (B).

Chapter 6

Conclusions and Future Work

Conclusions

The geological mechanisms involved in the formation of a deposit are in most cases transitional in nature, which yields contacts between domains that are diffuse or gradational. These soft boundaries are widespread in different types of deposits and their correct reproduction by geostatistical methods has a great impact on mine plan design, expected dilution and final mineral resources. The areas close to contacts are usually areas of higher uncertainty.

The estimation of a domain with a soft boundary implies that samples from either side of the boundary should be used in the estimation. A common practice is to include samples or previously estimated nodes from outside the domain within a certain distance. Whether kriging or simulation is used, the assumption that the samples or nodes outside the domain follow the same distribution and spatial model as the samples inside is often incorrect and leads to the corruption of the statistical parameters near the boundary.

One option explored in this work is the use of cokriging or cosimulation with a linear model of coregionalization (LMC) to capture the spatial correlation of the variable across a boundary between domains. In particular, the cross variogram of the variable of interest across a boundary cannot be calculated directly since the data sets are non-collocated. The alternative is to calculate the cross covariance. The extra time associated with fitting a linear model of coregionalization can be easily overcome by using semi-automatic fitting programs. This model allows the correlation of the grades across the boundaries to be captured through a legitimate spatial model of coregionalization, which can then be used to cokrige or cosimulate grades using data from adjacent rock types. This approach guarantees the correct reproduction of representative statistics of the individual geological units used for resource estimation.

This option has the advantage of improved resource estimation by reducing uncertainty in transitional zones near the boundaries. It reproduces data correlation across the boundary and has less smoothing of the estimates if kriging is the tool to obtain the resources.

The previous methodology, assumes that the variable is stationary in each domain, and therefore can be used to model a **global** correlation across a boundary. However, nature provides us with several examples where the behavior of our variable of interest is no longer stationary as it gets closer to a boundary. This work introduces a new technique for the estimation in the presence of **local** non-stationary soft boundaries.

This new technique corresponds to a non-stationary form of cokriging in the presence of geological boundaries. To apply this methodology the user must distinguish between stationary regions within each rock type and boundary zones where the statistical parameters such as the mean, variance or covariance are no longer constant. The non-stationary components of the mean and variance are optimized assuming a linear relationship with the distance to the boundary. This work has considered that the mean and variance increase towards the boundary. A decreasing mean near a boundary could be handled by a negative non-stationary mean, but this could lead to negative grade estimates. Decreasing variance near a boundary cannot be handled with this formalism. The non-stationary variance must be positive. A decreasing mean plus the proportional effect could decrease the variance, but it is a limitation of the presented methodology.

The correlation spatial model is also decomposed into stationary and non-stationary components of a linear model of coregionalization. The stationary component corresponds to the variograms of data within the stationary regions of each rock type. The non-stationary component corresponds to a relative standardized variogram scaled by the non-stationary standard deviation. The shape is given by the user and the range is optimized. All the optimization algorithms are a simplified version of the simulated annealing formalism, where only perturbations that minimize the objective function are accepted. With the stationary mean of each rock type plus its non-stationary linear model for each boundary zone and the linear model of coregionalization the estimation of unsampled locations is performed using a non-stationary form of simple cokriging. The data-to-data and the data-to-estimate covariances are calculated as the sum of the stationary and non-stationary components of the spatial model according to the location of the samples and estimate and whether or not they belong to a boundary zone. This new technique provides an appealing alternative to model the grade distribution in the presence of local non-stationary soft boundaries.

A set of FORTRAN programs were implemented for each of the steps described above and tested in some small examples as well as a large 3D application.

The estimates reproduce the non-stationary behavior of the mean of a reference distribution at the boundary zone. Cross validation results show that the result works much better than assuming a hard boundary. The exceptions are some blocks at the edges of the boundary zones where non-physical results are obtained due to unusual kriging weights. They seem to originate from the linear model of coregionalization; this must be explored in further detail.

The two methodologies presented in this work are theoretically robust. They can be easily applied in industry to improve resource estimation and simulation.

Future Work

An important task is to find the specific origin of non-physical covariance matrices that are the source of unusual kriging weights. There seems to be a discontinuity in the stationary and non-stationary components of the covariance model at the edges. When, for example, the location to be estimated is within a boundary zone and yet some samples used to estimate are outside the boundary zone. The covariance between the estimate and the sample turns out to be higher than the covariance of the sample to itself. Nevertheless, these are not the only configurations that lead to unusual estimates or negative kriging variances.

Another aspect that could be explored is the change in the correlation and/or anisotropy between adjacent domains within the boundary zone. This could be due to the occurrence of a secondary set of mineralized fractures at an angle to the principal mineralisation anisotropy.

The next step is the implementation of simulation in the presence of local non-stationary soft boundaries. It will be essential to have the correct kriging variance, which is the amount of missing variability that is added back to the estimate. The kriging variance has an additional non-stationary component, which has to be investigated in great detail to ensure the correct reproduction of the total variance at each location and the joint variability.

Other possible applications are indicator kriging and multivariate statistics. In the first, one can imagine soft boundaries between some or all categories or the estimation of different grade cutoffs in the presence of soft boundaries between geological domains; in both cases the challenge is to properly combine multiple

variograms in the corresponding linear model of coregionalization. A similar challenge would have to be addressed in the implementation of cokriging or cosimulation for multiple variables in the presence of a geological model with soft boundaries. In this case, the linear model of coregionalization should account for correlation between different variables as well as the spatial correlation of this set of variables across different rock types and the corresponding boundary zones.

References

C. V. Deutsch. *Geostatistical Reservoir Modeling*. Oxford University Press, New York, 2002.

C. V. Deutsch and S. Zanon. Ultimate SGSIM: Non-Stationary Sequential Gaussian Cosimulation by Rock Type. In *Centre For Computational Geostatistics*, volume 4, Edmonton, AB, 2002.

C.V. Deutsch and A.G. Journel. *GSLIB: Geostatistical Software Library: and User's Guide*. Oxford University Press, New York, 2nd Edition, 1998.

P. Goovaerts. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York, 1997.

A. G. Journel and C. J. Huijbregts. *Mining Geostatistics*. Academic Press, New York, 1978.

P. F. Larrondo, C. T. Neufeld and C. V. Deutsch. VARFIT: A Program for Semi-Automatic Variogram Modelling. In *Centre For Computational Geostatistics*, volume 5, Edmonton, AB, 2003.

T. M. Wawruch, C. V. Deutsch and J. A. McLennan. Geostatistical Analysis of Multiple Data Types that are not Available at the Same Locations. In *Centre For Computational Geostatistics*, volume 4, Edmonton, AB, 2002.